

Application Of Statistical Modeling To Interpret A Health System Crisis In Sri Lanka Due To COVID- 19

Indika Karunathilake¹, Mayuri Amarasiri¹, Anver Hamdani²

1. Faculty of Medicine, University of Colombo, Sri Lanka

2. Ministry of Health, Sri Lanka

Corresponding Author:

Prof. Indika Karunathilake

Head, WHO Collaborating Center for Medical Education

Faculty of Medicine

University of Colombo

Sri Lanka

e-mail: Indika@medarc.cmb.ac.lk

Abstract

This paper will discuss the application of statistic modeling to interpret a health system crisis in Sri Lanka due to COVID- 19.

A strong focus on the preventive approach and the contact tracing with the utilization of available resources in a rational manner describes Sri Lanka's response towards COVID- 19 prevention and mitigation. The early contact tracing, preemptive quarantining, isolation, and treatment were implemented as a concerted effort. This approach, proven efficient during the early phase of the pandemic, was sustainable when there was a rapid increase in the COVID- 19 patients since July 2021, exceeding the health system capacity.

The country's COVID- 19 situation during the period from 01st of August 2021 to 31st of October 2021 was taken into consideration. Variables used for analysis were; total number of cases,

recovered cases, comorbid and O₂ dependent patients, ICU patients, and deaths. The regression model was applied to analyze the data by using the EViews 12 (x64) software application.

The correlation coefficients of all the independent variables under consideration implies that they have a strong positive relationship with the number of deaths occurred during the said period. According to the computed multiple linear regression model, the number of positive cases and O₂ dependents have a positive relationship with the dependent variable. Further, the Durbin-Watson stat value of the model and multicollinearity test reflect that it is free from serial correlation thereby the model is fit. From the perspective of epidemiological control, these findings highlight the importance of keeping the number of cases within the limits of health system capacity.

Keywords: COVID- 19, Durbin- Watson statistic, Multiple Linear Regression, Multicollinearity

Introduction

COVID-19 was discovered in 2019 is an infectious disease caused by SARS-Cov-2. Since the outbreak of COVID- 19 pandemic, it has infected over 276,593,547 people with over 5,385,351 deaths worldwide as of 22nd December 2021. The World Health Organization (WHO) declared SARS- CoV-2 as a global pandemic on 11th of March 2020 as by that time it has spread over 100 countries within a matter of weeks. Since the World War II, the worst global crisis ever happened is COVID- 19 pandemic. At the very initial stage of the pandemic, the fatality rate of the COVID- 19 estimated by the WHO was approximately 2% though later the fatality rate was subjected to vary due to the mutations in coronavirus. In order to coordinate the public health response, the epidemic behavioral forecasting is particularly very supportive. This paper will discuss application of statistic modeling to interpret a health system crisis in Sri Lanka due to COVID- 19.

A strong focus on the preventive approach and the contact tracing with the utilization of available resources in a rational manner describes Sri Lanka's response towards COVID- 19 prevention and mitigation. The response to COVID- 19 in the country needs to be understood within the context of the health system existing in Sri Lanka. There is a strong and a well-established public health system in which the government provides health services free of charge in Sri Lanka (Karunathilake et al. 2020).

The early contact tracing, preemptive quarantining, isolation, and treatment can be recognized as an assertive and concerted effort in order to minimize the spread of the pandemic.

The salient features of the initial response of Sri Lanka to COVID- 19 pandemic were;

1. Early intervention by the government and the Ministry of Health
2. Extensive contact tracing
3. Early isolation of all the possible contacts pending confirmatory tests
4. Provision of well- organized quarantine facilities free of charge for the contacts with a high level of exposure
5. Isolation of the more vulnerable communities
6. Hospitalization of all the patients with positive PCR reports until their PCR becomes negative
7. Provision of free health care to all the patients

This approach was proven efficient during the early phase of the pandemic, although later it was evident that particular approach is not sustainable with the time as there was a rapid increase in the COVID- 19 patients since July 2021. The number of daily reported COVID- 19 positive patients exceeded 5000. This amounted to 220 cases per million population with a global rank of 50. The estimated actual number of new COVID- 19 new patients per day was around 6000 to 9000 during the peak of the pandemic. The available bed capacity for COVID- 19 patients in the country was increased to 34000 via the reallocations from the other units in the hospital by rationing the rest of the services to non- COVID patients as an immediate response in order to manage the crisis. Further, the prevention of deaths due to COVID-19 both at the hospital and at home prior to the admission became another challenging issue. The number of oxygen dependent patients and patient who needed ICU care was increasing beyond the health system capacity (Ministry of Health, 2021).

The available number of healthcare workers for the proper functioning of the health care system also decreased significantly due to the infection of COVID-19 and the prevailing workforce had to struggle with their physical and mental exhaustion amidst attempting to maintain the proper and smooth functioning of the healthcare services. Hence, Sri Lanka's healthcare system was reaching a tipping point with an unusual increase in the demand for the dwindling resources (Ministry of

Health 2021).

The dynamic nature in the pandemic highlighted the necessity for an accurate forecasting on how the pandemic will evolve. The Susceptible Infected Removed (SIR) epidemic model was proposed by Cooper, Mondal, and Anthonopulos (2020), based on the evolution of time of distinct populations and then monitored various variables which resulted the spread of the pandemic in countries and territories such as Australia, USA, China, India, South Korea and Italy. However, these models were not appropriate to interpret the crisis situation (Atkeson 2020; Bhaduri et al. 2020; Cooper et al. 2020)

Methodology

The publicly available information on COVID- 19 related information, and health outcomes consisting of the total number of cases, recovered cases, comorbid and O₂ dependent patients, ICU patients, and deaths were extracted from official sources; Health Promotion Bureau, Ministry of Health and the World Health Organization (WHO). The data existing from 01st of August 2021 to 31st October 2021 was considered in order to analyze, apply and interpret a statistical modeling for the health system crisis caused by the COVID- 19 pandemic.

Further the data on Sri Lankan healthcare capacity was also collected from the Annual Health Bulletins published by the Ministry of Health and Annual Reports of the Central Bank of Sri Lanka and it contained the information on the number of hospital beds, ICU beds, physicians and nurses per million of population.

The unit of analysis in this particular study is the country itself i.e. Sri Lanka. The regression model has been used in this study in order to analyze the data more comprehensively by using the EViews 12 (x64) software application. Thereby, the regression model which is applied to the Sri Lankan data can be utilized in order to predict or project the epidemic situation in a more pragmatic manner.

Results and Discussion

The following line graph shows how the total number of COVID- 19 positive cases and deaths occurred during the period under consideration in Sri Lanka

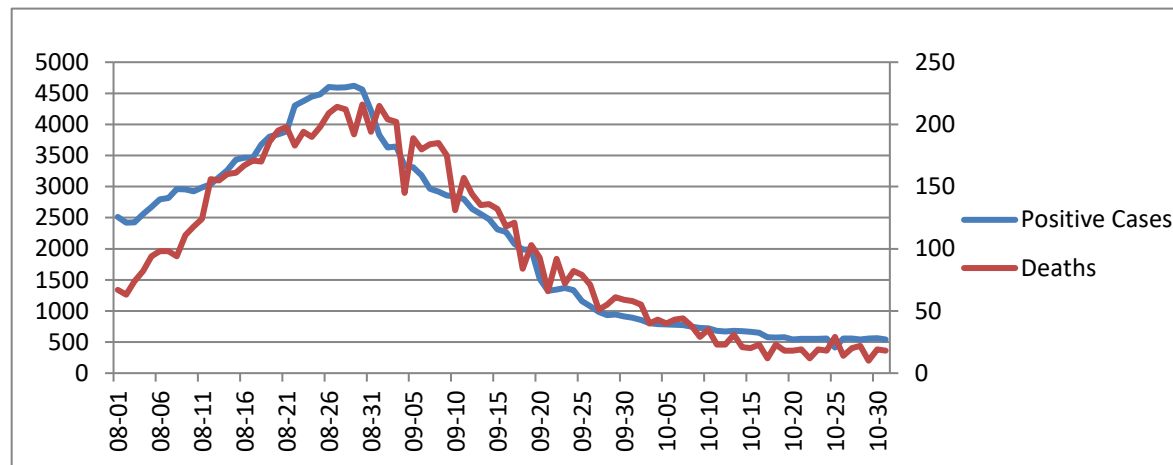


Figure 1: Total no. of confirmed positive cases and deaths in Sri Lanka (01.08.2021- 31.10.2021)

Firstly, it is necessary to investigate the descriptive statistics of the data set in order to get the micro level understanding of the study as it provides a brief and precise summary of the samples and measures under consideration.

Table 1: Descriptive Statistics

	DEATHS	POSITIVE_CASES	O2_DEPENDENT	TOTAL_ICU_PTS	COMORBID
Mean	101.1087	2135.217	522.9022	148.0761	2679.272
Median	92.50000	2292.500	524.0000	176.5000	2946.500
Maximum	216.0000	4621.000	1033.000	196.0000	4647.000
Minimum	10.00000	412.0000	91.00000	69.00000	634.0000
Std. Dev.	68.14923	1380.137	315.0490	47.65749	1451.858
Skewness	0.235391	0.275754	0.011022	-0.616484	-0.182180
Kurtosis	1.593946	1.678521	1.552893	1.614770	1.397074
Jarque-Bera	8.428060	7.860129	8.029314	13.18312	10.35817
Probability	0.014787	0.019642	0.018049	0.001372	0.005633
Sum	9302.000	196440.0	48107.00	13623.00	246493.0
Sum Sq. Dev.	422632.9	1.73E+08	9032286.	206682.5	1.92E+08
Observations	92	92	92	92	92

Source: Survey Data, 2021

Skewness is identified as a measure of asymmetry of the distribution in the series around its mean value which is computed as follow:

$$S = \frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - \bar{y}}{\hat{\sigma}} \right)^3$$

Where $\hat{\sigma}$ is the estimator in order to measure the standard deviation, which is based on the

biased estimator for the variance. In a normal distribution, the skewness of a symmetric distribution is zero. When there is a positive skewness, it is identified that the distribution has a long right tail while the negative skewness value denoted that the distribution has a long left tail. Hence, in this study, it is recognized that no. of deaths, positive cases and O₂ dependents have a positive skewness while the number of ICU patients, and comorbid have a negative skewness in the distribution.

Kurtosis is the measurement of the peakedness or the flatness of the distribution of the series and it is computed as:

$$K = \frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - \bar{y}}{\hat{\sigma}} \right)^4$$

in which $\hat{\sigma}$ is based on the biased estimator for the variance. In a normal distribution, the kurtosis is 3. If the distribution is peaked (leptokurtic) relative to normal, then the kurtosis will exceed 3. On contrast, if the distribution is flat relative to the normal i.e. platykurtic, then the kurtosis will be less than 3. In this particular computed data series, it is identified that the kurtosis values are all less than 3 which means that the data series is platykurtic.

Jarque- Brea is a test statistic which is used for testing in order to identify whether the data is normally distributed. This Jarque- Bera test can be identified as goodness- of- fit test in order to recognize whether the sample data have the skewness and kurtosis matching a normal distribution.

The test statistic JB can be defined as:

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4}(K - 3)^2 \right)$$

where the n is the no. observations or degrees of freedom, sample skewness is denoted with the S and K is the sample kurtosis. Hence, according to Jarque-Bera statistic, all these variable data is not normally distributed as they do not exceed the probability value of 0.05.

The analysis on correlation coefficients can be used in order to measure the strength of the linear

relationship between the two variables. If the correlation coefficient is greater than zero, it implies a positive relationship where as if the correlation coefficient value is less than zero then it indicates a negative relationship. This can be further analyzed using the scatterplot diagram in order to understand the nature and the direction of the relationship between the variables.

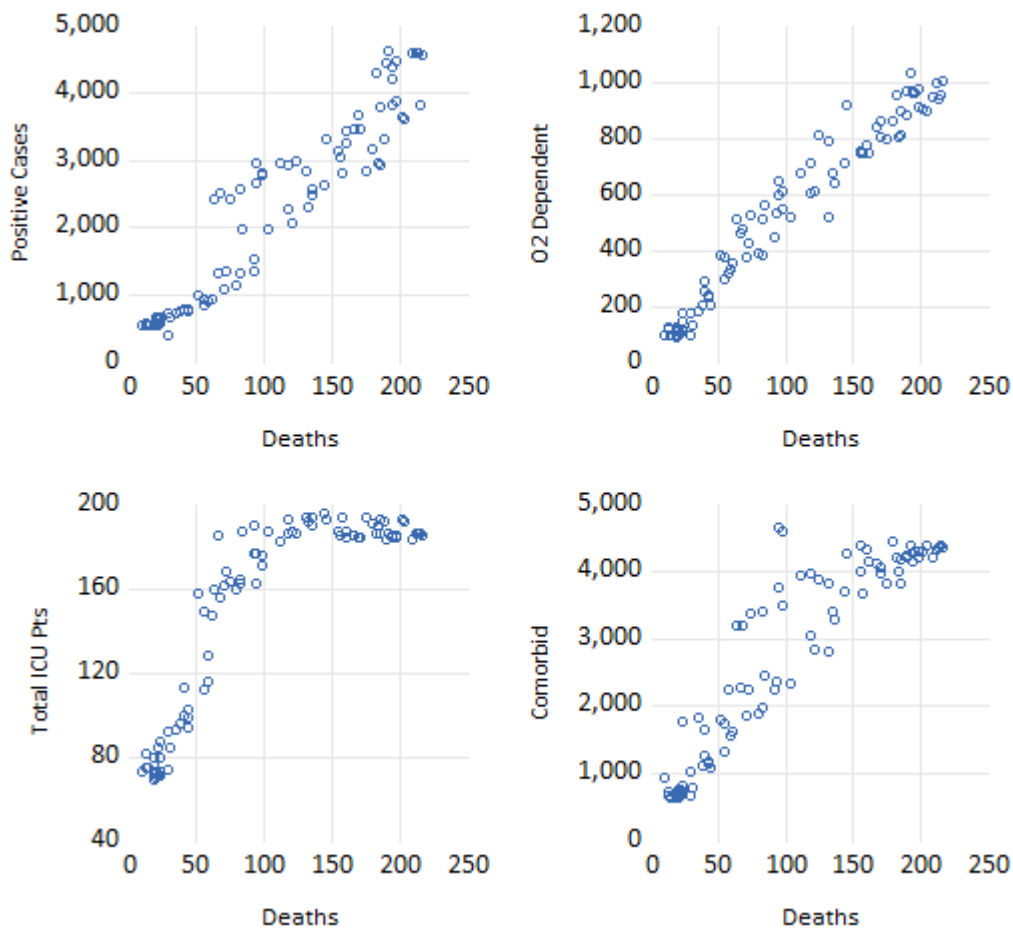


Figure 2: The scatter diagrams for the variables under consideration.

The summary of the correlation coefficients for the variables under consideration can be shown as follow.

Table 2: Summary table of the correlation coefficients

	Positive cases	O ₂ Dependents	No. of patients in the ICU	Comorbid
Deaths	0.953976	0.975982	0.861985	0.920705

Source: Survey Data, 2021

According to the above computed correlation coefficient table, it is identified that the relationship

of each of the independent variables have a high strong positive relationship with the dependent variable i.e. no. of deaths as the coefficient correlation values are positive and higher than 0.75.

Multiple linear regression attempts to model the relationship between two or more independent or explanatory variables and one dependent or response variable by fitting a linear equation to the observed data. In the data series, every value of the independent variable x is associated with a value of the dependent variable y . The model is used in order to estimate the relationship between two or more independent variables and one dependent variable.

Table 3: Multiple Linear Regression

Dependent Variable: DEATHS

Method: Least Squares

Sample: 8/01/2021 10/31/2021

Included observations: 92

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.206937	7.204834	0.167518	0.8674
POSITIVE_CASES	0.004963	0.006043	0.821291	0.4137
O2_DEPENDENT	0.251852	0.031288	8.049587	0.0000
TOTAL_ICU_PTS	-0.076823	0.098791	-0.777628	0.4389
COMORBID	-0.011575	0.004398	-2.632108	0.0100
R-squared	0.958855	Mean dependent var	101.1087	
Adjusted R-squared	0.956963	S.D. dependent var	68.14923	
S.E. of regression	14.13781	Akaike info criterion	8.188398	
Sum squared resid	17389.37	Schwarz criterion	8.325452	
Log likelihood	-371.6663	Hannan-Quinn criter.	8.243714	
F-statistic	506.8642	Durbin-Watson stat	1.548062	
Prob(F-statistic)	0.000000			

Source: Survey Data, 2021

According to the computed multiple linear regression model, the coefficient values capture the estimates for the independent variables i.e. no. of positive cases, O₂ dependents, comorbid, and ICU patients and the intercept is denoted with c . The positive signs of the coefficients imply that the variables of no. of positive cases and O₂ dependents have a positive relationship while comorbid and ICU patients have a negative relationship with the dependent variable i. e. no. of deaths.

The t- statistic value for number of O₂ dependents implies that it is sufficient evidence against the

null hypothesis as the t- stat is above 2 whereas t- statistics values for no. of positive cases, comorbid and ICU patients imply that those are not sufficient evidence to reject null hypothesis as the t- stat is less than 2.

In the computed multiple regression model, it shows that the number of O² dependents, and comorbid patients' has a p- value which is less than 0.05 which implies a higher level of confidence in rejecting the null hypothesis while the p- value for the number of positive cases and ICU patients are higher than p- value of 0.05 which means that the null hypothesis can't be rejected for those particular independent variables.

The R- squared value of 0.958855 gives the variation in the independent variables that are explained by positive cases. The higher the R², the better the model and more predictive power the variables have.

The F- statistic shows whether the explanatory variables are significant in explaining the outcome variable, deaths. In this multiple linear regression model the F- statistic value is 506.8642 depicts that higher the F- stat it is better for the model. Further, it can be stated that the model is significant as the Prob (F- statistic) is less than the 0.05 level.

The Durbin- Watson stat value of 1.548062 reflects that it is free from serial correlation.

The lower the value of these criteria, the better the model is. From this computed multiple linear regression model, it is identified that Akaike Info Criterion (AIC) figure of 8.188398 is the lowest of the three and therefore indicates that it is the best model to adopt. The AIC is an estimator for the prediction error and thereby relative quality of statistical models for the given set of data as the AIC estimates the relative amount of loss of the information by a particular model. If the model loses a less amount of information then it implies that the model is of a higher quality.

In the normality test, the term normality refers to the normal distribution or sometimes it is known as the bell shaped curve or the Gaussian distribution, which can be identified as a specific

statistical distribution.

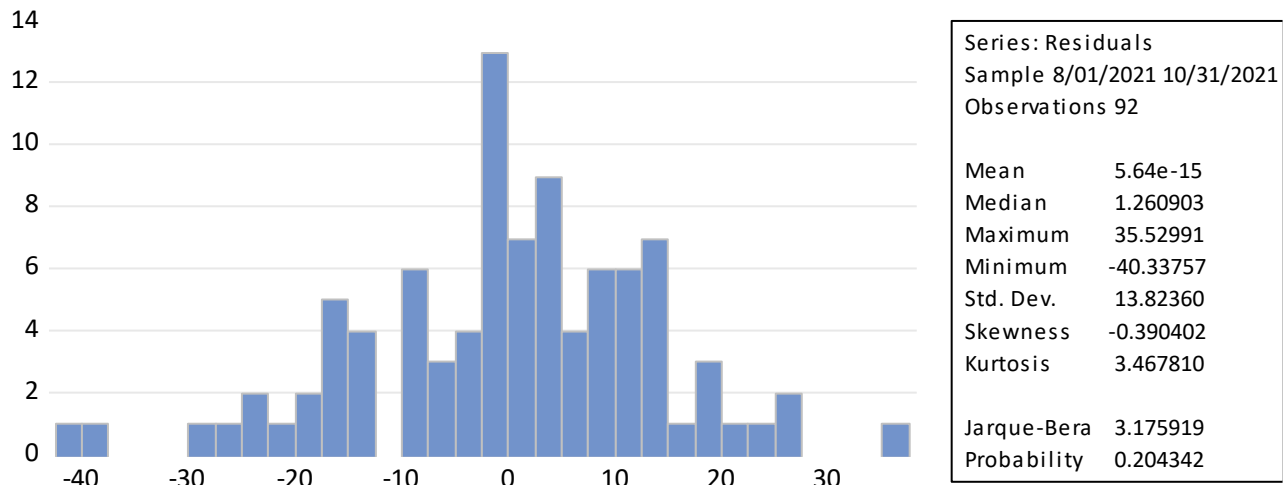


Figure 3: Normality test

According to this normality test result, it is identified that the p-value i.e. 0.204342 is larger than the significance level of 0.05. Thereby the decision is to fail to reject the null hypothesis as there is no adequate evidence to conclude that the data do not follow a normal distribution.

Table 4: Multicollinearity test

Variance Inflation Factors

Sample: 8/01/2021 10/31/2021

Included observations: 92

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	51.90964	23.89303	NA
POSITIVE_CASES	3.65E-05	108.2932	31.66617
O2_DEPENDENT	0.000979	167.4353	44.23610
TOTAL_ICU_PTS	0.009760	108.5898	10.09190
COMORBID	1.93E-05	82.46255	18.56029

Source: Survey Data, 2021

In the multicollinearity test, the Variance Inflation Factors (VIFs) is a method in which measures the collinearity between the regressors in the equation. The VIFs show the level of variance of a coefficient estimate of a regressor has been inflated because of the collinearity with the other regressors. It is simply calculated by dividing the variance of a coefficient by the variance of that coefficient had other regressors not been included in the equation.

Hence according to the VIF values and tolerance values (Centered VIF value) it can be stated that

the model is not free from multicollinearity issue as the uncentered VIF values are greater than 10 and centered VIF values are greater than 0.2.

Conclusion

The main purpose of this study was to apply a statistical model in order to investigate the impact of COVID- 19 on the health system in Sri Lanka. Therefore, the country's COVID- 19 situation during the period from 01st of August 2021 to 31st of October 2021 was taken into consideration. The analysis was done based on the regression model. The correlation coefficients of all the independent variables under consideration implies that they have a strong positive relationship with the number of deaths occurred during the said period. According to the computed multiple linear regression model, the number of positive cases and O₂ dependents have a positive relationship while comorbid and ICU patients have a negative relationship with the dependent variable. Further the Durbin- Watson stat value of the model and multicollinearity test reflect that it is free from serial correlation thereby the model is fit. From the perspective of epidemiological control, these findings highlight the importance of keeping the number of cases within the limits of health system capacity.

References

Atkeson, A., 2020. *How deadly is COVID-19? Understanding the difficulties with estimation of its fatality rate* (No. w26965). National Bureau of Economic Research.

Bhaduri, R., Kundu, R., Purkayastha, S., Kleinsasser, M., Beesley, L.J. and Mukherjee, B., 2020. Extending the susceptible-exposed-infected-removed (SEIR) model to handle the high false negative rate and symptom-based administration of Covid-19 diagnostic tests: SEIR-fansy. *Medrxiv*.

Central Bank of Sri Lanka. (2021). Annual Report [online].

Available from: <https://www.cbsl.gov.lk/en/publications/economic-and-financial-reports/annual-reports> [accessed 23 December 2021].

Cooper, I., Mondal, A. and Antonopoulos, C.G., 2020. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons & Fractals*, 139, p.110057.

Department of Census and Statistics. (2021). Population [online].

Available from: <http://www.statistics.gov.lk/Population/StaticallInformation/CPH2011>

[accessed 22 December 2021].

Karunathilake, I., Mudiyanse, R., Witharana, C. and Peiris, M., (2020). The rational approach for controlling the spread of COVID-19 in Sri Lanka. *Ceylon Medical Journal*, 65(3), pp.72–74.

Ministry of Health. (2021). Epidemiology Unit [online].

Available from: <https://www.epid.gov.lk/web/> [accessed 24 December 2021].

Ministry of Health. (2021). Annual Health Bulletin [online].

Available from: http://www.health.gov.lk/moh_final/english/others.php?pid=110

[accessed 23 December 2021].

The World Bank. (2021). World Bank Open Data [online].

Available from: <https://data.worldbank.org/> [accessed 22 December 2021].

World Health Organization. (2021). WHO Sri Lanka [online].

Available from: <https://www.who.int/srilanka> [accessed 22 December 2021].