ORIGINAL ARTICLE e-ISSN: 2349-0659 p-ISSN; 2350-0964

Modeling the Determinants of COVID-19 Mortalities in South-East Asia

V. Rema[1]*, K. Sikdar[2]

Abstract

Modeling has been used extensively to predict coronavirus disease 2019 cases across different countries. This research provides a zone-wise macro perspective in understanding the determinants of pandemic deaths in select Asian countries. Correlations are established with variables impacting daily deaths, and panel regression analysis is carried out. Comorbid conditions, especially the cardiovascular death rates in the population and existence of diabetes condition do have a significant impact on pandemic death rates. Age is also a determinant of death rates. Countries with lesser comorbid rates in their population with increased access to hospital beds have significantly seen lower death rates. The study aids in decision making, providing a macro perspective on how specific global regions have responded with regard to pandemic deaths.

Keywords: Coronavirus disease 2019, Feasible generalized least squares model, Pandemic mortalities, Panel data regression, South-East Asia

Asian Pac. J. Health Sci., (2022); DOI: 10.21276/apjhs.2022.9.1.37

Introduction

The outbreak of the coronavirus disease 2019 (COVID-19) was reported to the World Health Organization (WHO) in December 2019, first observed in Wuhan City, China. The world has been battling the pandemic since then. The global countries are classified into six zones by the WHO, namely, Africa, America, South-East Asia, Europe, Eastern Mediterranean, and Western Pacific to monitor and implement strategic plans in each of the zones when the necessity emerges. Globally, as on June 13, 2021; 175,306,598 confirmed cases of COVID-19, including 3,792,777 deaths are reported.[1] The case-fatality ratio, that is, the ratio of the deaths among the infected global population is 0.022 (around 2%), which clearly indicates the trend that recovery rates are over 95%. While most normal infected individuals recover, it shows that fatalities are associated with certain high-risk population sub-groups. The presence of pre-conditions, certain additional ailments, age, care facilities available in the location may be specific reasons pushing individuals into that undesirable zone of 2%.

Older patients (age >65 years) with existence of co-morbid conditions were seen to have severe impacts with intensive-unit care requirements and worse levels of prognosis when affected by COVID-19. A robust electronic review of literature by Sanyaolu *et al*.[2] and data obtained from Research Square, a meta-analysis of the 2019 novel coronavirus by Paudel[3] showed hypertension, cardio vascular diseases, diabetes, respiratory illness such as chronic obstructive pulmonary disease (COPD) to be among the major comorbid conditions among patients affected with COVID-19. Smoking is one of the biggest risk factors for COPD. There are many studies in literature that have used time series, Susceptible, Exposed, Infectious and Recovered (SEIR) models etc. for analyzing the pandemic cases. Reiner *et al*.,[4] in their study used COVID-19 case and mortality data from February to September 2020 and forecasted using a deterministic SEIR compartmental framework. Their study further reiterated the use of masks and adoption of social distancing which quantitatively reflected the control of spread. Another work by Olsen *et al*.[5] used National Family Health Survey data for 2015-16, Census data of 2011 and COVID 19 deaths data up to June 2020 and analyzed death trends in India through a hierarchical multilevel model. Age, obesity, existence of comorbid conditions, and smoking habits were among the key conditions showing higher death rates due to COVID-19 infection. Giang *et al*.[6] assessed WHOs and World Bank's Development indicators for a short period from January to March 2020 for factors affecting pandemic deaths. As per their panel analysis; availability of hospital beds, medical staff, age, air travel, quarantine and social distancing measures were found crucial in affecting mortality rates. Daniyal *et al*.[7] used data from the National Institute of Health of Pakistan from February to August 2020 related to COVID 19 deaths and other demographic, socioeconomic metrics. They compared linear, logarithmic, and quadratic regression models, and found the quadratic model to be the best fit, for modeling COVID-19 deaths in Pakistan.

Their study showed death rates due to COVID were higher among men than women, by attributing it to heart diseases, smoking habits, etc., being higher in men, making them more susceptible to deaths. Panel data modeling is shown to be a validated statistical method and a useful addition to the epidemiological toolkit for understanding the patterns of pandemic progression. Oehmke *et al*.[8] used dynamic panel data models with the Arellano-Bond estimator using the generalized method of moments framework for modeling the pandemic

[1]Department of Business Analytics and Data Science, Ramaiah Institute of Management, Bengaluru, Karnataka, India

[2]Department of Mathematics, BMS Institute of Technology and Management, Bengaluru, Karnataka, India

Corresponding Author: V. Rema, Department of Business Analytics and Data Science, Ramaiah Institute of Management, Bengaluru, Karnataka, India. E-mail: venkit.rema@gmail.com

How to cite this article: Rema V, Sikdar K. Modeling the Determinants of COVID-19 Mortalities in South-East Asia. Asian Pac. J. Health Sci., 2022;9(1):158-163.

Source of support: Nil

Conflicts of interest: None.

Received: 18/09/21 **Revised:** 17/10/21 **Accepted:** 02/12/21

situation in the United States. The model is suitable as it corrects for autocorrelation and heteroscedasticity, usually evident in macro panels. Their study was based on the metrics of speed- the number of cases per day, acceleration- the movement of number of cases per day, whether increasing or decreasing, and the jerk-the change in acceleration. Staszkiewicz *et al*.[9] carried out panel analysis by applying random effects model on data from December 2019 to March 2020, using the pandemic statistics data from the EU Open Data Portal, the World Bank Open Data, the market data from Stooq and weekly data from Google Trends, to investigate pandemic progression. Among the several variables assessed, the country-specific variables were age dependency, number of physicians and nurses per 1000 of population, the fraction of the population with the handwashing facilities, the density of the population, the level of the pollution, the number of the available hospital beds per 1000 of the population, etc. Their study pointed out to severity being different across continents and also analyzed the impact of time invariant variables on death progression. The variables for the present study are selected based on the literature; prominent variables that are likely to have an impact on the COVID-19 death rates are considered. Studies capturing a panel approach with a macro perspective on a group of countries are limited. This study focuses on assessing some of the key determinants of daily deaths and analysis of pandemic death scenario among select countries in Asia.

Materials and Methods

Data Collection

Data are extracted from a public repository[10] with eight variables for select eight countries in the Asian region. The data are panel data as it is across countries which are the cross-sectional units, over a period of time. Panel data have more degrees of freedom and sample variability than cross-sectional data which provides efficient estimates.[11] The variables considered for the study are daily deaths (daily-death); positivity rate, that is, share of the COVID-19 tests turning out positive, given as a 7-day rolling average (pos-rate); country-wise population median age (med-age); cardiovascular death rates, that is, annual number of deaths per 100,000 people due to cardiovascular diseases (CVD-rate); diabetes prevalence rates, that is, percent of the population aged 20–79 affected with diabetes as per latest UN report (diab-rate); percent of population involved in smoking (smoke-rate); hospital beds available per 1000 of the population (hosp-beds) and human development index (HDI) of the country. The daily data are taken for a year from March 2020 to February 2021.

Methodology

Descriptive analysis is carried out. Correlations are computed to assess the variables associated with daily deaths. Variables that lead to multicollinearity (Variance inflation factor [VIF] >10)[12] have been removed so as to present the actual relationship between daily deaths modeled as a function of its predictors. Data are divided into training and test data set in 70:30 ratio. Heterogenous panel regression models of fixed effects, random effects and feasible generalized least squares (FGLS) are applied to factor the variations in the panel data. The models are fit to the data and forecast accuracy is measured using root mean square error (RMSE) and compared between training and test data. R-software is used for data analysis.

Results and Discussion

Across the panel of eight countries in South-East Asian zone- Bangladesh, India, Indonesia, Myanmar, Nepal, South Korea, Sri Lanka and Thailand; the average daily deaths seen over a year was 83, with the highest being 40% of tested cases turning out to be positive on a 7-day rolling average basis. Among the countries considered; India, Indonesia, Bangladesh and Myanmar have seen higher daily deaths compared to the others [Figure 1]. Higher median age, higher hospital beds per 1000 population and higher HDI (>0.75) is seen in South Korea, Sri Lanka and Thailand; with <200 annual cardiovascular related deaths (per 100,000 population) in these countries. India and Sri Lanka have over 10% of its population with diabetes prevalence. On an average, 40% of the population across the countries reflects smoking habits [Table 1].

Out of the considered variables, those with higher VIF (i.e., VIF range of 7–10 are removed to control multicollinearity between predictors. The considered predictors are positivity rate, median age, cardiovascular death rate, diabetes rate and availability of hospital beds per 1000 population, which are all significantly related to daily deaths with $P < 0.05$. Further, positivity rate, cardiovascular death rate, and diabetes rate are positively correlated with daily deaths. Median age and availability of hospital beds are negatively correlated with daily deaths [Table 2]. India, Indonesia, Bangladesh and Myanmar with higher positivity rates and cardiovascular, diabetes rates have shown higher death rates; while South Korea, Sri Lanka and Thailand have shown lesser daily deaths comparatively, with the overall population showing lesser percentage of comorbid conditions. South Korea, Sri Lanka and Thailand with higher older population (i.e., higher median age) have shown lesser deaths. Further, these countries with more access to hospital beds have also shown significantly fewer deaths.

The variables are logarithmically scaled and the Augmented-Dickey Fuller test is run to show series stationarity for daily deaths at 10% level of significance ($P = 0.08$).[13] On the trained data set, the heterogenous panel regression model of fixed effects and random

Table 1: Panel descriptive statistics

Variable	*Minimum*	*Mean*	*Maximum*	*Standard deviation*
Daily-death	0	82.6	2003	227.1
Pos-rate	0	0.1	0.4	0.1
Med-age	25	32.1	43.4	6.1
CVD-rate	86	222.4	342.9	84.9
Diab-rate	4.6	7.7	10.7	1.9
Smoke-rate	20.6	40.1	76.1	15.4
hosp-beds	0.3	2.7	12.3	3.8
HDI	0.583	0.707	0.916	0.106

HDI: Human development index, CVD: Cardiovascular deaths

Table 2: Correlation metrics with daily deaths

Variable	*r*	*P-value*	*VIF*
Pos-rate	0.13	2.758×10^{-8}	2.09
Med-age	–0.23	$<2.2\times10^{-16}$	6.44
CVD-rate	0.28	$<2.2\times10^{-16}$	6.43
Diab-rate	0.40	$<2.2\times10^{-16}$	1.09
Hosp-beds	–0.20	$<2.2\times10^{-16}$	2.97

CVD: Cardiovascular deaths

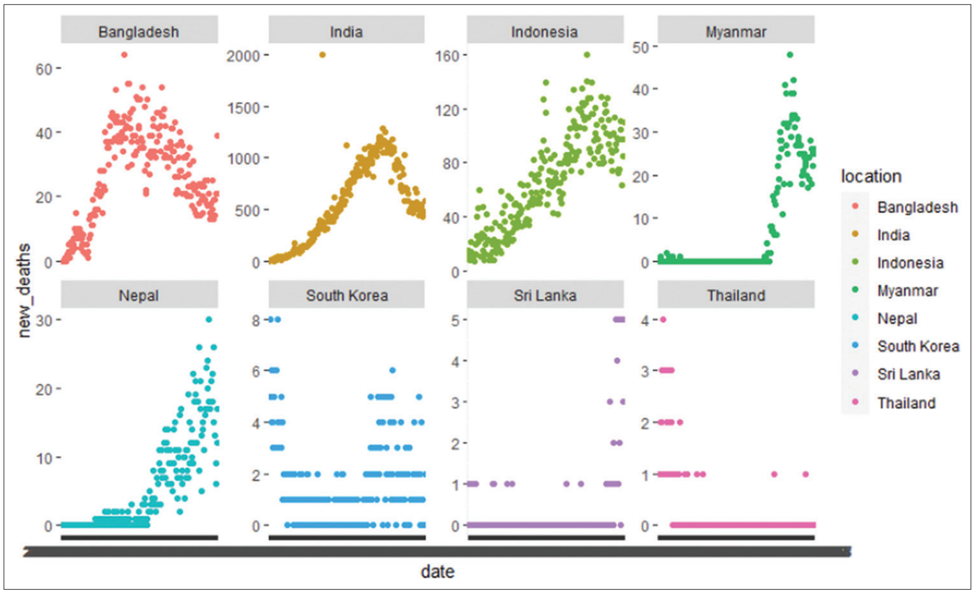

Figure 1: Country-wise daily deaths

effects are applied.[14] The one-way fixed effects model, with "within estimator" is first applied to capture country-specific effects. This is followed by the one-way Random effects model (Equation 1), where the error term accounts for individual-heterogeneity.

$$y_{it} = \mu + \beta_1 X_{1it} + \beta_2 X_{2it} + ... + \beta_k X_{kit} + (\alpha_i + u_{it}), \quad u_{it} \sim IID(0, \sigma_u^2) \qquad (1)$$

y_{it} represents daily deaths where i = Country (i = 1–8), t = days; μ is the constant intercept; X_{kit} is the k^{th} log transformed independent variable for the i^{th} country and t^{th} time; β_k is the regression coefficient for the k^{th} independent variable; α_i is the cross-sectional effect; u_{it} is the error term, u_{it}~IID (0, σ_u^2). The output in Table 3 shows the model is overall significant with $P < 2.22 \times 10^{-16}$; positivity rate, median age and cardiovascular death rates are significant in predicting daily deaths (at 10% level of significance). The model overall has 23% explanatory power, with adjusted r^2 being 0.2244, which is realistic since the possible number of variables that can have an impact on the deaths are significantly very large and could be individual-specific, varying from case to case. Here, country-wise daily deaths are obtained. The model gives a realistic predictive power with a macro perspective and confirms a significant role of age, positivity rate and comorbid conditions on pandemic deaths.

The Hausman test iterates the null hypothesis of consistency of random effects over fixed effects model and is a test to identify the model of best fit between the two.[15] The Hausman test yields a $P = 0.4363$ (fail to reject the null hypothesis), which suggests that random effects model is preferred over the fixed effects model and therefore, model output of random effects is only shown. The model assumptions of homoscedasticity and existence of auto correlations of residuals are tested using the Breusch and Pagan (BP) test,[16] and Wooldridge test,[17] respectively. The null hypotheses of the BP test state the existence of homoscedasticity and that of the Wooldridge test, non-existence of serial correlation of residuals. Both the tests yield significant results, that is, $P < 2.2 \times 10^{-16}$, implying existence of both heteroscedasticity and serial correlation. Real time large macro panels do witness this frequently and are best handled by considering robust standard errors to get consistent estimates, which is factored in Table 3, where the robust standard errors are considered.[18] The residuals are symmetric, showing normality [Figure 2]. The panel residual graphs country-wise is plotted [Figure 3] and residuals have a mean of -8.86×10^{-17} (~0) with standard deviation of 0.8454.

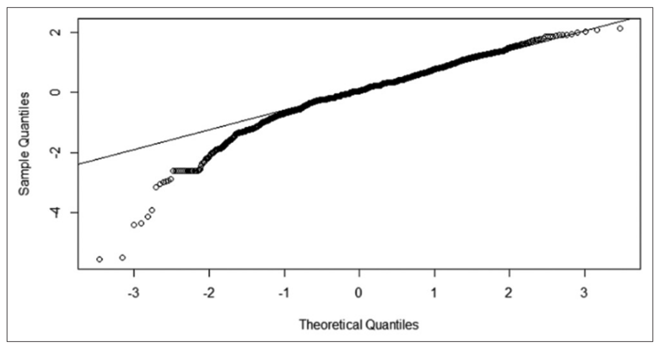

Figure 2: Quantile-Quantile plot of residuals

Table 3: Random effects model output summary

Variable	*Estimate*	*Robust standard error*	*t-value*	*P-value*
Intercept	−66.3103	28.8949	−2.2949	0.0218
Pos-rate	0.4080	0.0793	5.1443	2.96×10^{-7}
Med-age	13.2623	7.1399	1.8575	0.0634
CVD-rate	4.2380	1.1031	3.8419	0.0001
Diab-rate	0.8879	1.7574	0.5052	0.6135
Hosp-beds	−1.1950	1.0698	−1.1171	0.2641

CVD: Cardiovascular deaths

In the presence of heteroskedasticity and serial correlations, a model framework more suitable is the FGLS which provides consistent estimates with Prais-Winsten correction and Panel Corrected Standard Errors (PCSE).[19-21] The output [Table 4] shows the model is overall significant with $P = 0$; positivity rate, median age, cardiovascular death rates, and availability of hospital beds are significant in predicting daily deaths (at 10% level of significance).

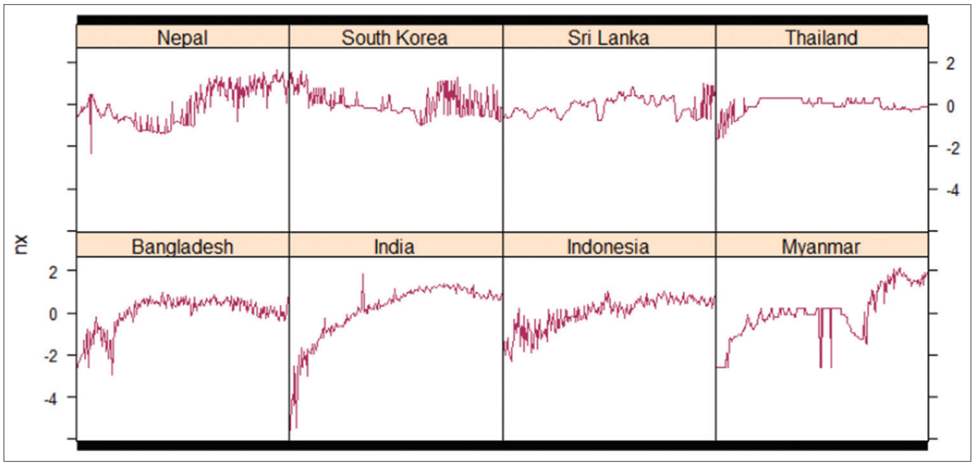

Figure 3: Panel residual plots

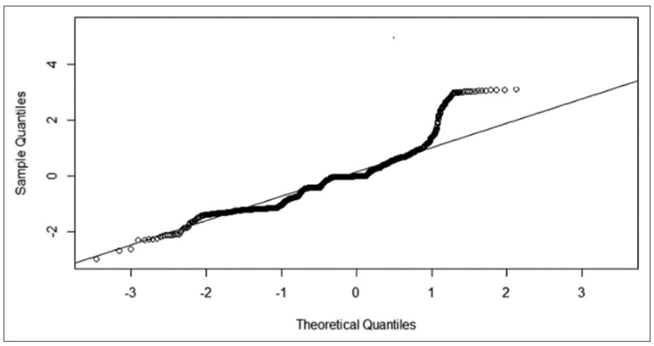

Figure 4: Quantile-Quantile plot of residuals

Table 4: FGLS-PCSE model output summary for trained data

Variable	*Estimate*	*Panel corrected standard error*	*t-value*	*P-value*
Intercept	−44.8128	7.5082	−5.969	2.85×10^{-9}
Pos-rate	0.0631	0.0176	3.582	0.0004
Med-age	5.7578	1.9161	3.005	0.0027
CVD-rate	4.9267	0.3483	14.145	$<2 \times 10^{-16}$
Diab-rate	0.3447	0.3067	1.124	0.2612
Hosp-beds	0.4598	0.2766	1.663	0.0966

FGLS-PCSE: Feasible generalized least squares-panel corrected standard errors, CVD: Cardiovascular deaths

The model overall has 25% explanatory power, with adjusted r^2 being 0.2542.

The model assumptions are tested. The residuals are fairly symmetric [Figure 4] indicating normality. The residuals have a mean of 0.0773 (~0) with standard deviation 1.7345, satisfying the assumptions. It can be seen that the standard errors are much smaller in FGLS model compared to the Random effects model, which is essential for the estimates to be consistent. Hence, the model validation is carried out for the test data by applying the FGLS model with panel corrected standard errors.

The FGLS model is overall significant with 18% of variations in daily deaths explained by significant predictors of positivity rate, median age, cardiovascular death rates, and availability of hospital beds. These variables are highly significant in predicting daily deaths ($P < 0.01$) [Table 5]. The result aligns with what was seen for the training set. The residual plots are generated country-wise

Table 5: FGLS-PCSE model output summary for test data

Variable	*Estimate*	*Panel corrected standard error*	*t-value*	*P-value*
Intercept	−174.367	28.3482	−6.151	1.21×10^{-9}
Pos-rate	0.4614	0.1679	2.748	0.0061
Med-age	42.6497	7.3234	5.824	8.32×10^{-9}
CVD-rate	6.5057	0.8251	7.885	1.03×10^{-14}
Diab-rate	−0.6754	0.7329	−0.922	0.3570
Hosp-beds	−4.4082	0.9499	−4.641	4.05×10^{-6}

FGLS-PCSE: Feasible generalized least squares-panel corrected standard errors, CVD: Cardiovascular deaths

Table 6: Forecast accuracy measure-RMSE

Country	*RMSE train*	*RMSE test*
Bangladesh	1.724	2.217
India	51.929	4.124
Indonesia	1.473	0.714
Myanmar	4.659	2.894
Nepal	2.733	17.745
South Korea	1.663	3.777
Sri Lanka	0.090	0.709
Thailand	0.884	0.001
Overall mean	8.144	4.023

RMSE: Root mean square error

based on the FGLS model [Figure 5] with residual mean of −0.6238 (~0) and standard deviation 2.5162. The fits were generated post retransformation of data and forecast accuracy was checked by evaluation of the RMSE.[22] The RMSE is comparable for training and test data [Table 6].

Conclusions

The countries in the Asian region with better access to healthcare and lower overall population comorbid metrics have responded well to the pandemic with reasonable death rates than the others. In this study, South Korea, Thailand and Sri Lanka come into this category, although they have in general higher median aged population, compared to countries such as India, Bangladesh, and Nepal. Hence to achieve lower death rates, it is vital to provide better access to the healthcare system and ensure additional precautions for individuals with comorbid conditions. A micro picture on district-wise, town-wise modeling will provide a perspective on handling pandemic deaths at a particular location. However, this study is

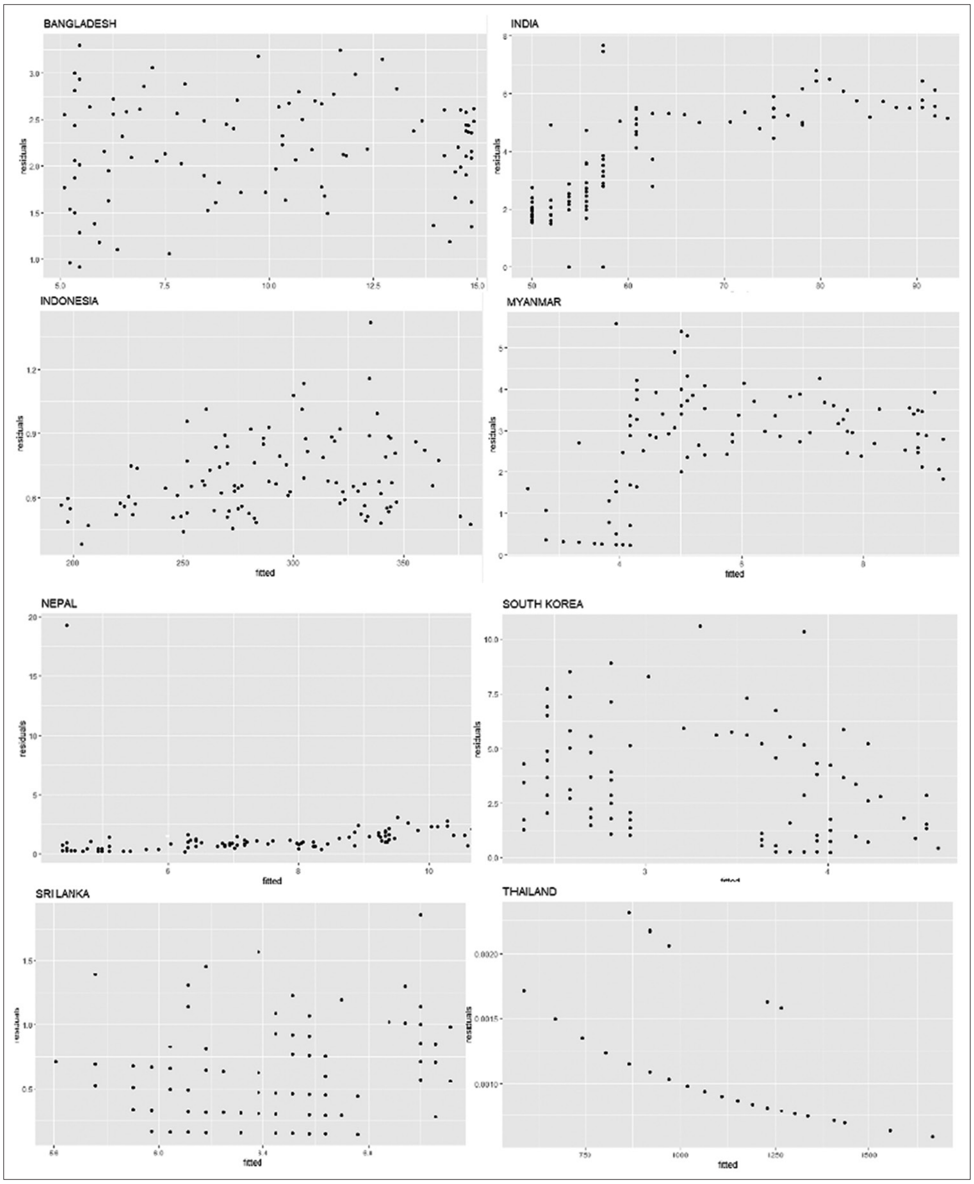

Figure 5: Country-wise residual plots

significant as it provides a broad overview of how pandemic deaths are affected in the South-East Asian member states classified by the WHO. The WHO mobilizes resources, logistics and makes strategic plans for its various zone classifications, looking at how the member states in the zone have responded to the pandemic. The aid for decision making, providing a bird's eye view is the focus in this study. There is future scope for research by considering a greater number of predictors that can affect the pandemic death trends, including response to the vaccinations.

Acknowledgments

The first author thankfully acknowledges the R and D Centre, support and encouragement provided by the Management and staff of the Department of Mathematics, BMS Institute of Technology and Management and Ramaiah Institute of Management, Bengaluru.

Data Availability

Data source is public repository OurWorldInData.org and World Bank Open Data latest year country-metrics. Relevant data have been extracted from this source for the study undertaken.

Copyright and Permission Statement

We confirm that the materials included in this paper do not violate copyright laws. Where relevant, appropriate permissions have been obtained from the original copyright holder(s). All original sources have been appropriately acknowledged and/or referenced.

References

1. WHO. WHO Coronavirus (COVID-19) Dashboard. Geneva: WHO; 2021. Available from: https://www.covid19.who.int [Last accessed on 2021 Jun 13].
2. Sanyaolu A, Okorie C, Marinkovic A, Patidar R, Younis K, Desai P, *et al.*

Comorbidity and its impact on patients with COVID-19. SN Compr Clin Med 2020;2:1069–76.
3. Paudel SS. A Meta-Analysis of 2019 Novel Coronavirus Patient Clinical Characteristics and Comorbidities; 2020. Available from: https://www.researchsquare.com/article/rs-21831/v1 [Last accessed on 2020 May 01].
4. COVID-19 Forecasting Team. COVID-19 Scenarios for the United States, medRxiv; 2020. p. 94-105.
5. Olsen W, Bera M, Dubey A, Kim J, Wiśniowski A, Yadav P. Hierarchical modelling of COVID-19 death risk in India in the early phase of the pandemic. Eur J Dev Res 2020;32:1476-503.
6. Giang TL, Vo DT, Vuong QH. COVID-19: A relook at healthcare systems and aged populations. Sustainability 2020;12:4200.
7. Daniyal M, Ogundokun RO, Abid K, Khan MD, Ogundokun OE. Predictive modeling of COVID-19 death cases in Pakistan. Infect Dis Model 2020;5:897-904.
8. Oehmke JF, Moss CB, Singh LN, Oehmke TB, Post LA. Dynamic panel surveillance of COVID-19 transmission in the United States to inform health policy: Observational statistical study. J Med Internet Res 2020;22:e21955.
9. Staszkiewicz P, Chomiak-Orsa I, Staszkiewicz I. Dynamics of the COVID-19 contagion and mortality: Country factors, social media, and market response evidence from a global panel analysis. Ieee Access 2020;8:106009-22.
10. Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. Coronavirus Pandemic (COVID-19); 2021. Available from: https://www.ourworldindata.org/coronavirus [Last accessed on 2021 Mar 01].
11. Hsiao C, Mountain DC, Illman KH. A Bayesian integration of end-use metering and conditional-demand analysis. J Bus Econ Stat 1995;13:315-26.
12. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. Berlin, Germany: Springer Science & Business Media, 2011.
13. Cowpertwait PS, Metcalfe AV. Introductory Time Series with R. New York: Springer; 2010.
14. Baltagi BH. Econometric Analysis of Panel Data. Chichester: John Wiley & Sons; 2008.
15. Hausman JA. Specification tests in econometrics. Econometrica 1978;46:1251-71.
16. Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. Econometrica 1979;47:1287-94.
17. Drukker DM. Testing for serial correlation in linear panel-data models. Stata J 2003;3:168-77.
18. Millo G. Robust standard error estimators for panel models: A unifying approach. J Stat Softw 2017;82:1-27.
19. Gujarati DN, Porter DC. Basic Econometrics. New Delhi, London: Tata MCGraw; 2004.
20. Baltagi BH. Forecasting with panel data. In: Center for Policy Research Working Paper No. 91; 2007. Available from: https://www.surface.syr.edu/cpr/74 [Last accessed on 2020 Jun 01].
21. Bailey D, Katz JN. Implementing panel corrected standard errors in R: The pcse package. J Stat Softw 2011;42:1-11.
22. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature. Geosci Model Dev 2014;7:1247-50.