Article │ Received 15 July 2023; Accepted 2 August 2023; Published 11 December 2023
https://doi.org/10.55092/pcs2023020032

Computational assessment of COVID-19 time series using ARIMA models: comparative analysis of new cases, vaccination rates, booster doses, and mortality trends

Mogana Darshini Ganggayah[1,*], Ruben Rajandaran[2], Nurul Aqilah Mohd Nor[3] and Nurhasniza Idham Abu Hasan[4]

[1] Department of Econometrics and Business Statistics, School of Business, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500 Subang Jaya, Selangor
[2] Napier Technologies Limited, 51 Level 12, Menara Binjai, Jalan, Binjai, 50450 Kuala Lumpur, Malaysia
[3] School of Pharmacy, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500 Subang Jaya, Selangor
[4] Department of Statistics, Faculty of Computer & Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia

* Correspondence author; Email: moganadarshini.ganggayah@monash.edu.

Abstract: Coronavirus Disease, COVID-19 has had a crucial effect on both worldwide health and economies, underscoring the importance of examining the time-based trends of essential metrics. This research presents a comprehensive time series analysis of COVID-19 data in Malaysia starting March 2020 until May 2023, specifically focusing on new cases, vaccination rates, booster doses, mortality rates and ICU patients' trends during pre- and post-vaccination. The study employs a time series ARIMA model to investigate the interrelationships and dynamics among these variables during the pandemic. The study utilizes a publicly available real-time dataset from Our World in Data platform, providing a detailed and comparative analysis of the different variables. The study findings reveal significant correlations between the variables, shedding light on the effectiveness of vaccination campaigns, the impact of booster doses, and their influence on reducing new cases and mortality rates. The mortality rates and ICU patients count reduce gradually with the implementation of the vaccination program. The quantity of new Covid19 cases reduces gradually with the increasing number of fully-vaccinated individuals followed by booster doses. This research enhances our comprehension of the ever-changing COVID-19 situation and guides decision-making based on solid evidence to effectively reduce the future impact of COVID-19.

Keywords: time series analysis; vaccination rate; ARIMA model; booster doses; new cases; new deaths; COVID-19; ICU patients; Malaysia

1. Introduction

COVID-19 has now surpassed 768 million cases, accompanied by a staggering toll of over 6.9 million fatalities [1]. Furthermore, it has been observed that a minimum of 10% of these cases lead to the manifestation of persistent long-term COVID symptoms [2,3] numerous patients encounter many symptoms affecting various organ systems [4]. The figure is most likely to be higher due to numerous undocumented cases. Mortality in COVID-19 patients who need close monitoring, mechanical ventilation, and over the duration of the continuing pandemic, the supply of organs evolved tremendously. [5,6]. COVID-19 vaccinations are expedited on a large scale globally, and as of July 2023, about 13.5 billion COVID-19 vaccinations and 2.8 billion booster doses have been administered [1]. Vaccines have been proven to prevent cases and their benefits are well established [7,8], including the vulnerable groups [9,10].

Following the government's implementation of vaccine mandates in Malaysia, approximately 84.9% of the total population, equivalent to 27.5 million out of 32.4 million, have now received the COVID-19 vaccine. These mandates have led to restrictions on entry into public spaces, especially government facilities, for individuals who have not completed their vaccination [11]. The rate of vaccination acceptance among Malaysians is low due to various reasons [12–17], only half of the population have taken booster doses and these are mainly taken by the people who live in urban areas for example Putrajaya, Kuala Lumpur and Selangor [18,19].

While earlier research primarily concentrated on identifying factors influencing vaccination acceptance rates [12,20–22], there has been limited investigation into the patterns of COVID-19 cases and mortality before, during, and after vaccination. There is a pressing need for practical tools that can bolster the monitoring along with tracking of COVID-19 while also improving the effectiveness of medical systems practically [23,24]. Studying the effects of vaccination patterns may enhance the vaccination rate and aid the policy-makers in making timely decisions, which helps reduce the COVID-19 mortality rate.

COVID-19 has significantly impacted countries worldwide, necessitating the exploration of effective strategies to combat the pandemic [25]. Analyzing time series data offers valuable insights into the dynamics of COVID-19, enhancing our comprehension and predictive capabilities for future trends [26], to look into the links between COVID-19 new occurance and various related variables such as total vaccinations, booster doses, mortality and number of ICU (intensive care unit) patients in Malaysia.

2. Methods

2.1. Collection of data and preprocessing

The real-time COVID-19 data was obtained from the Our World in Data platform [27], which contained information for multiple countries. We filtered the dataset to include only the data from Malaysia. The raw data consisted of 67 variables and 295,919 rows. We selected a subset of variables, including new cases, vaccination rate, booster doses, new deaths, and the number of ICU patients, for the time series analysis. To study the impact of vaccination, we divided the data into two periods: before vaccination (pre-vaccination) and after vaccination (post vaccination). The before vaccination period ranged from 24th March 2020 to 23rd February 2021. The after vaccination period ranged from 24th February 2021 until 27th May 2023, where the booster dose counts started on 1st September 2021.

2.2. Time series analysis

In this study, ARIMA model, which combines elements of autoregression (AR), differencing (I), and moving average (MA) has been applied to analyses the Covid19 data. The ARIMA model enables the examination of past values, trends, and seasonality to forecast future observations accurately. Figure 1 shows the methodology of the ARIMA model utilised in this research. Throughout this research, the time series analysis was performed using Python programming.

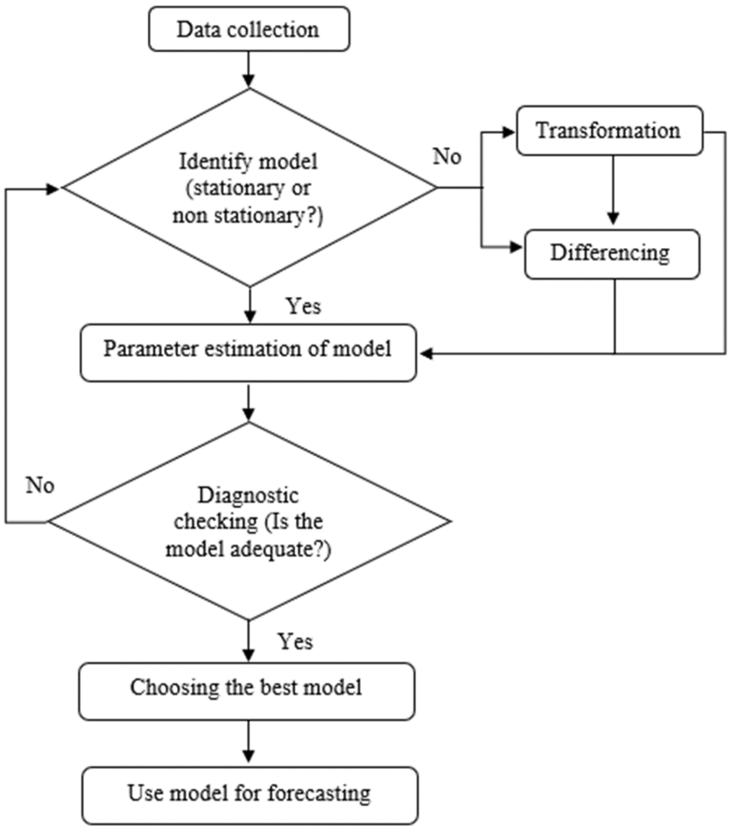

Figure 1. ARIMA model time series forecasting methodology.

The formulation is as follows for the ARIMA (p, d, q) model:

AR(p): The autoregressive element factors in the connection between the present both current and previous observations. It presupposes that the time series' current value is linearly dependent on the past p observations, weighted by corresponding coefficients. Mathematically, it is represented as:

$$Xt = c + \varphi 1 * Xt - 1 + \varphi 2 * Xt - 2 + \ldots + \varphi p * Xt - p + \varepsilon t \tag{1}$$

Here, The variables Xt and c are the time series at time t, 1 to p are the autoregressive coefficients, and t is the error term.

I(d): The element of differentiation aims to make the time series stationary by taking differences between consecutive observations. This step is crucial as it helps remove trends and seasonality that may exist in the data. The differencing operation is applied d times until the series becomes stationary.

MA(q): The impact of previous forecast errors on the present observation is captured by the moving average component. It presupposes that the time series' current value is linearly dependent on the past q forecast errors. Mathematically, it is represented as:

$$Xt = c + \theta 1 * \varepsilon t - 1 + \theta 2 * \varepsilon t - 2 + \ldots + \theta q * \varepsilon t - q + \varepsilon t \quad (2)$$

Here, θ1 to θq are the coefficients of the moving average, and εt is the term of residual error.

The ARIMA model in this study was typically implemented in four steps:

During this initial phase, we ascertained the orders for the moving average (q) and autoregressive (p) components by evaluating the time series partial autocorrelation function (PACF) charts and data's autocorrelation function (ACF).

The model parameters (φ and θ coefficients) used least squares estimation.

We scrutinized the model's residuals to confirm that they adhered to specific assumptions, including mean of zero with normal distribution and consistent variance. We also conducted diagnostic tests to evaluate the randomness of these residuals.

Once the model has been validated, it was used to make future predictions by generating forecasts based on the estimated coefficients and the available data.

The time series started by comparing new cases overall and number of vaccinations. Following the commencement of vaccination on 24 February 2021, the data during the post-vaccination period was analyzed to observe any shifts or patterns resulting from the vaccination campaign. Additionally, the impact of booster doses was investigated, which were introduced on 1st September 2021. By considering this variable, the study focused to evaluate the possible implications of booster doses on the selected COVID-19 indicators.

3. Results

The time series analysis using the ARIMA model allowed the research team to uncover valuable insights regarding the relationship between COVID-19 variables in Malaysia. The results demonstrated notable changes in new cases, vaccination rate, booster doses, new deaths, and the number of ICU patients during before vaccination and after vaccination periods. By analyzing the impact of booster doses, additional fluctuations were observed in the selected COVID-19 indicators.

Time series analysis

To obtain the values of q and p, the partial autocorrelation function (PACF) and autocorrelation function (ACF) were studied in ARIMA model. Figure 2 shows the ACF and PACF plots. Lag 15 is used as p value and Lag 8 is used as q value.

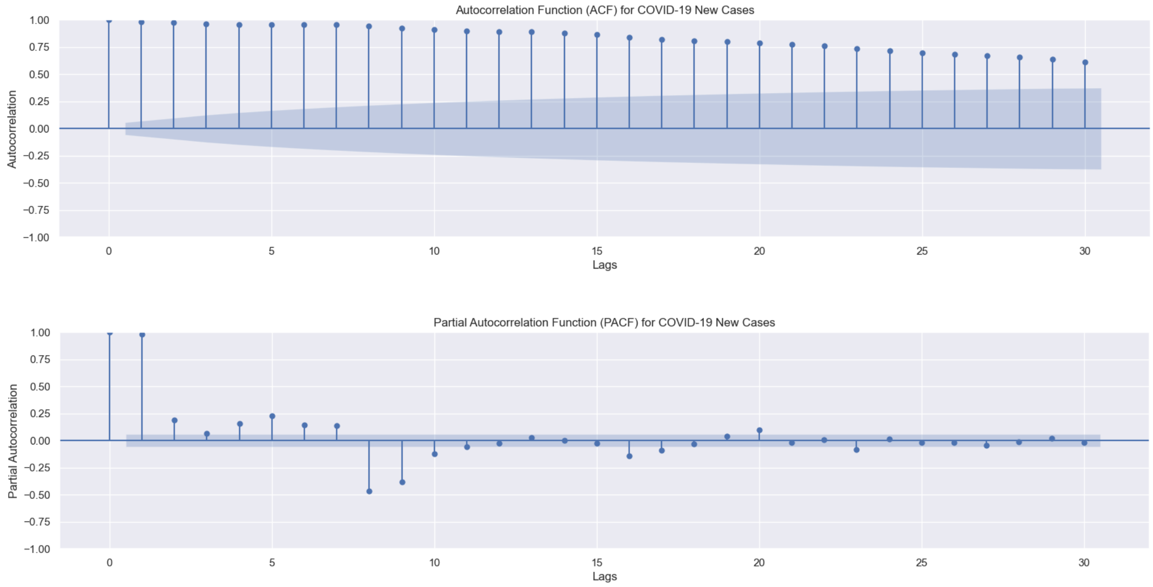

Figure 2. The for autocorrelation function (ACF) and partial autocorrelation function (PACF), respectively Covid19 New Cases.

Four comparative time series analyses were performed in this study. Figure 3 illustrates a study of time series for the cumulative count of new COVID-19 cases before and after the vaccination duration, emphasizing the implications of the immunization process. The x-axis's time scale goes from the pandemic's start to the most recent data point, while the y-axis denotes the number of new cases reported daily. Prior to the initiation of the vaccination process, the time series exhibits a consistent upward trend, indicating a rapid increase in the overall quantity of new COVID-19 cases. This period, referred to as the pre-vaccination phase, is marked by an alarming surge in infections and often corresponds to the initial stages of the pandemic. After approximately one year of the vaccination process (December 2021), a noticeable change in the trend becomes evident. The overall rate of new COVID-19 cases starts to drop steadily, suggesting the potential impact of widespread vaccination efforts. This phase, referred to as the post-vaccination phase, signifies the period when a significant portion of the population has received vaccination against the virus.

Figure 4 presents a comparative analysis of new COVID-19 cases over time, specifically focusing on three distinct periods: pre-vaccination, post-vaccination, and post-booster dose implementation. Following the implementation of widespread vaccination efforts, the post-vaccination phase demonstrates a notable decline in how many new COVID-19 cases there are. The curve shows a downward trend, suggesting that vaccinations have effectively reduced the overall transmission rate and contributed to limiting the spread of the virus. The post-booster dose implementation phase is characterized by a flattening curve, indicating a stable and reduced number of new cases. This decline and subsequent stabilization in cases suggest that booster doses have played a crucial role in reinforcing the immune response and providing additional protection against COVID-19.

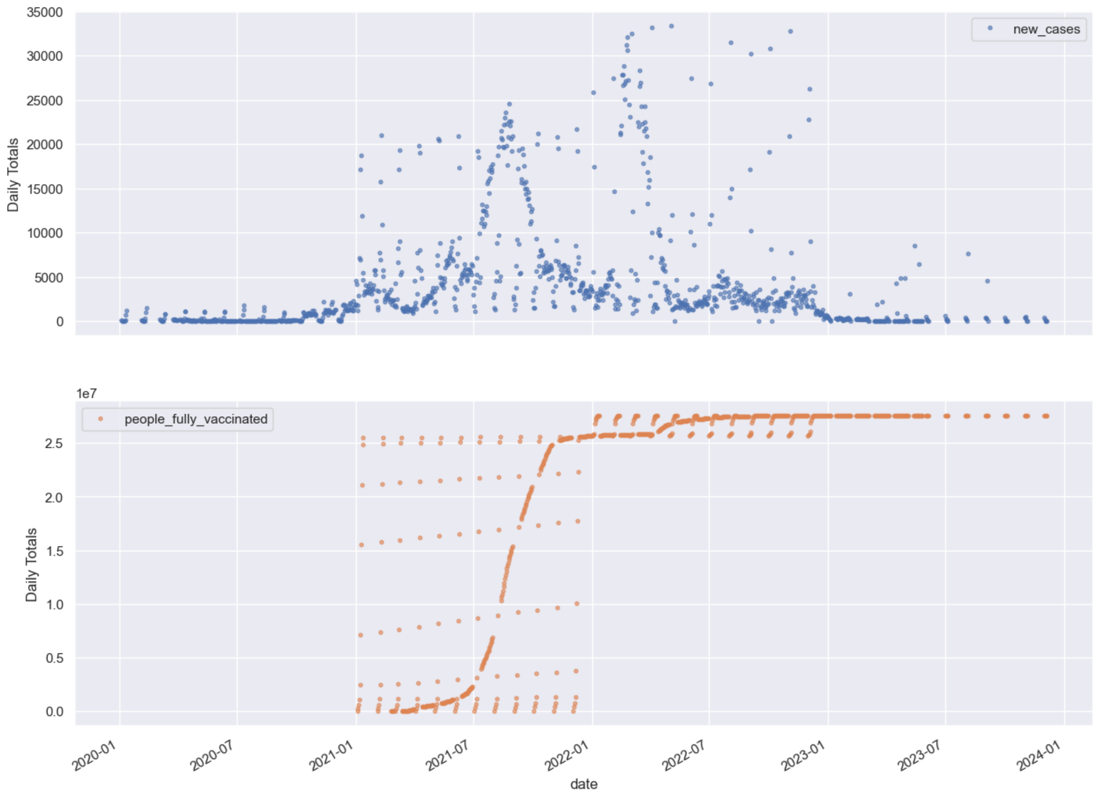

Figure 3. New Covid19 cases *versus* total number of vaccinations pre- (before January 2021) and post-vaccination (from January 2021).

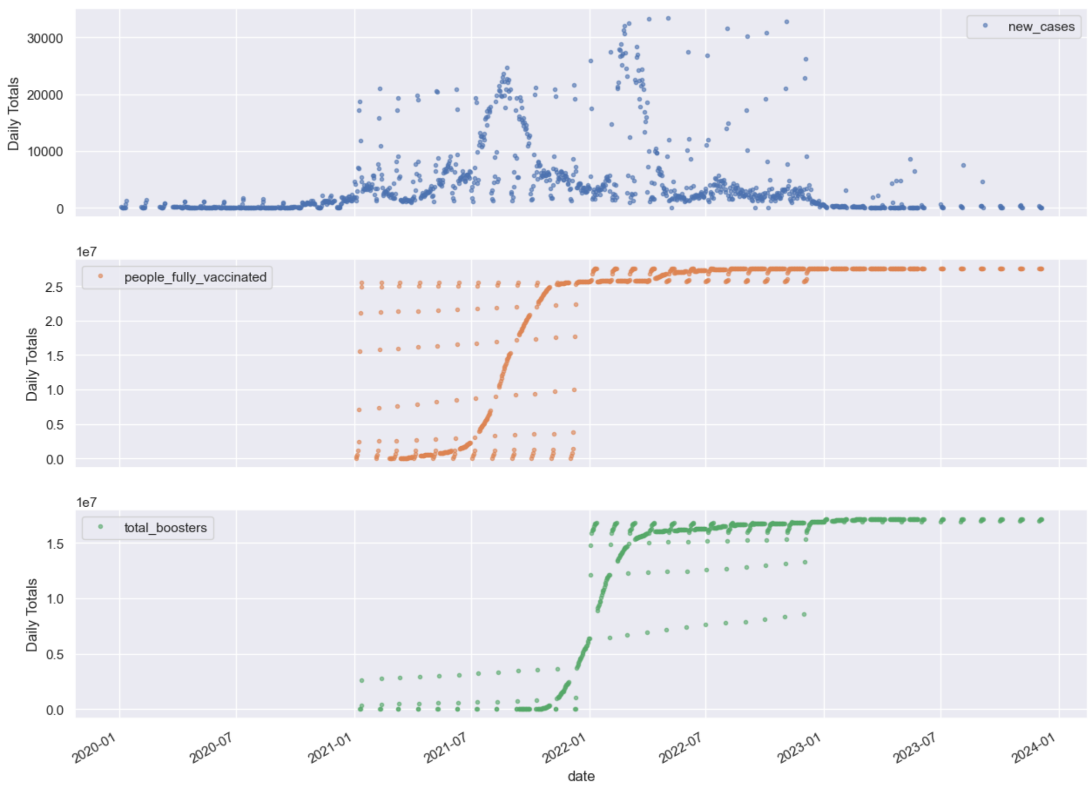

Figure 4. New Covid19 cases *versus* total number of vaccinations and booster doses.

Figure 5 presents a time series analysis that contrasts the overall number of immunizations with the quantity of new COVID-19 fatalities and booster doses administered. The curve representing new COVID-19 deaths demonstrates the impact of the pandemic on mortality rates over the specified time period. It initially exhibits a rising trend during the pandemic's early phases, highlighting the devastating toll of the virus on human lives. However, as vaccination efforts progress, the curve shows a decline, indicating a decrease in the quantity of fatalities associated with COVID-19. Additionally, the count of booster doses administered is represented on the same plot, providing insights into the administration of supplementary doses to individuals who have previously been vaccinated, showing lower counts in mortality rate.

Figure 5. New deaths, total number of vaccinations and booster doses.

Figure 6 presents a time series analysis that compares the number of total ICU (Intensive Care Unit) patients with the counts of total vaccinations and total booster doses administered. The curve representing total ICU patients illustrates the burden on healthcare systems due to severe cases requiring intensive care. Initially, this curve exhibits an upward trend, indicating a growing number of individuals requiring critical care as the pandemic progresses. However, with the initiation of vaccination efforts, the curve starts to gradually decrease, suggesting a potential reduction in severe cases and ICU admissions. Moreover, the count of booster doses administered is represented on the same plot, indicating the administration of additional doses to individuals who have already been vaccinated. The number of ICU patients becomes consistent after the implementation of booster doses.

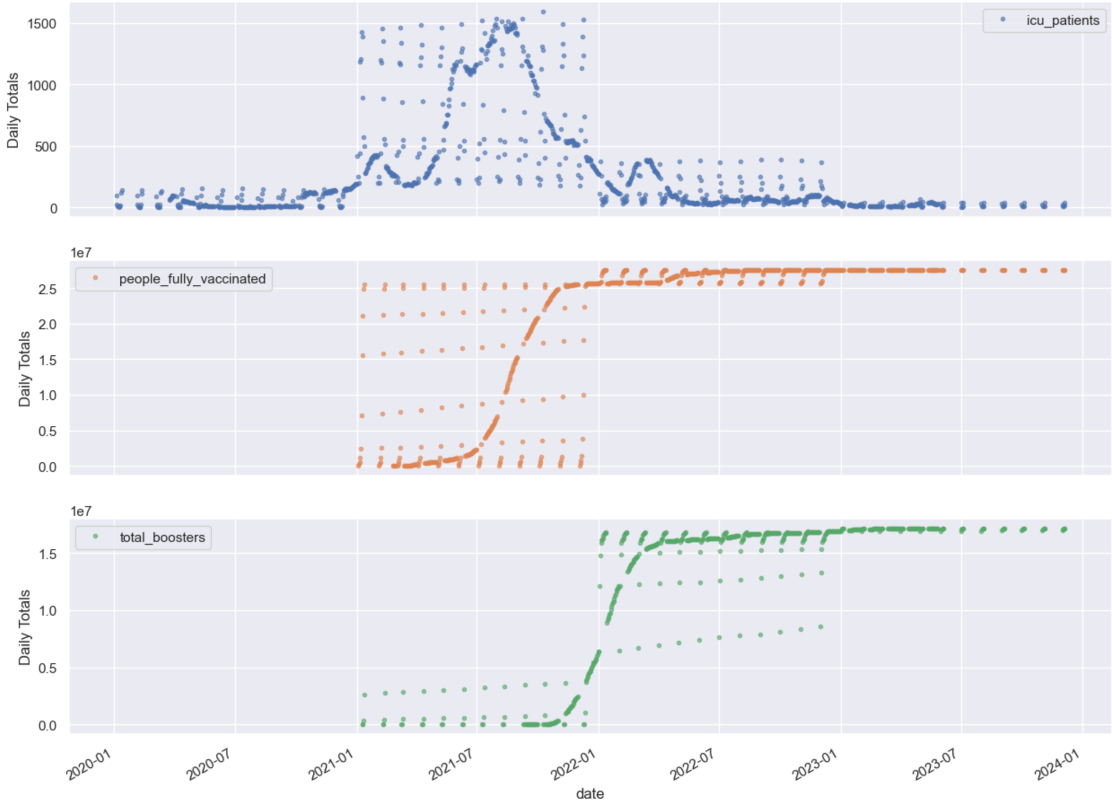

Figure 6. New ICU patients, number of vaccinations and booster doses.

4. Discussion

The time series analysis using the ARIMA model allowed us to uncover valuable insights regarding the relationship between COVID-19 variables in Malaysia. The results demonstrated notable changes in new cases, vaccination rate, booster doses, new deaths, and the number of ICU patients during before vaccination and after vaccination periods. By analyzing the impact of booster doses, we observed additional fluctuations in the selected COVID-19 indicators.

The research outcomes align with earlier studies, which have consistently shown that a robust vaccination effort can lower the mortality rates associated with COVID-19. One study recommends that a swift vaccination rollout is the most effective policy and strategy reaction to a pandemic crisis because it lowers the death toll and the number of infected people. Additionally, vaccination has significantly reduced COVID-19 hospitalizations, with fewer cases needing intensive care unit (ICU)-level mechanical ventilation treatment [7]. Another study demonstrates the high effectiveness of two vaccine doses in preventing ICU admissions or the need for mechanical ventilation to support life.

The decline in new COVID-19 cases noted during the post-vaccination phase in this study underscores the success of the vaccination campaign in curbing the virus's spread. It illustrates how vaccinations can effectively lower the overall transmission rate while also providing protection against severe illness and hospitalization. The time series analysis offers a graphical depiction of vaccinations' favorable influence on managing the COVID-19 pandemic. This underscores the significance of widespread immunization initiatives in reducing occurrence of new cases, as well as underscores the pivotal role of vaccination as a vital strategy in addressing public health emergencies.

The comparative analysis of new COVID-19 cases across three periods (pre and post-vaccination, post-booster dose) highlights the impact of vaccination and booster dose implementation in mitigating the spread of the virus. It underscores the significance of both initial vaccinations and subsequent booster doses in reducing the number of new cases and potentially preventing severe illness and hospitalization. The time series analysis presented in this study enables the comparison of the mortality rate and the number of total ICU patients with the counts of vaccinations and booster doses. It allows for the observation of potential relationships between the progress of vaccination campaigns and the impact on ICU admissions.

The results from this study highlight how the gradual reduction in ICU patients may coincide with the rollout of vaccination efforts and the subsequent implementation of booster dose programs which could assist the policy makers for future prevention and to create more awareness campaigns about vaccination for the public.

5. Conclusion

This research employed the ARIMA model to perform a comprehensive study of the COVID-19 time series data in Malaysia. By investigating the relationship between COVID-19 variables, such as new cases, vaccination rate, booster doses, new deaths, and the number of ICU patients, valuable insights were gained into the impact of vaccination on the pandemic dynamics. These results aid in a better comprehension of the efficiency of vaccination campaigns and the importance of booster doses in managing the COVID-19 crisis. As demonstrated, the ARIMA model serves as a guide for this investigation, potent instrument for time series analysis and forecasting in the context of pandemics and can help decision-makers control and reduce the spread of COVID-19 by providing them with information they need.

Acknowledgement

The authors would want to extend my sincere gratitude to all the people and organisations who helped make this research study possible.

Conflicts of interests

The authors declare no conflict of interests.

References

[1] WHO COVID-19 Dashboard. Available: https://covid19.who.int/data (accessed 13 July 2023).
[2] Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. *Nat. Rev. Microbiol.* 2023, 21(3):133–146.
[3] Ballering AV, Zon SKR van, Hartman TC olde, Rosmalen JGM. Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. *The Lancet* 2022, 400(10350):452–461.

[4] Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, *et al.* Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* 2021, 38: 101019.
[5] Chandel A, Leazer S, Alcover KC, Farley J, Berk J, Jayne C, *et al.* Intensive Care and Organ Support Related Mortality in Patients With COVID-19: A Systematic Review and Meta-Analysis. *Crit. Care Explor.* 2023, 5(3): e0876.
[6] Auld SC, Harrington KRV, Adelman MW, Robichaux CJ, Overton EC, *et al.* Trends in ICU Mortality From Coronavirus Disease 2019: A Tale of Three Surges. *Crit. Care Med.* 2022, 50(2):245–255.
[7] Hoxha I, Agahi R, Bimbashi A, Aliu M, Raka L, *et al.* Higher COVID-19 Vaccination Rates Are Associated with Lower COVID-19 Mortality: A Global Analysis. *Vaccines* 2023, 11(1):74.
[8] Thomas SJ, Moreira ED, Kitchin N, Absalon J, Gurtman A, *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine through 6 Months. *N. Engl. J. Med.* 2021, 385(19):1761–1773.
[9] Cavanna L, Citterio C, Toscani I. COVID-19 Vaccines in Cancer Patients. Seropositivity and Safety. Systematic Review and Meta-Analysis. *Vaccines* 2021, 9(9):1048.
[10] Powers AC, Aronoff DM, Eckel RH. COVID-19 vaccine prioritisation for type 1 and type 2 diabetes. *Lancet Diabetes Endocrinol.* 2021, 9(3):140–141.
[11] Tin DTP. Covid-19 Cases And Deaths By Age (Jan 2020 – Feb 3, 2023). Available: https://codeblue.galencentre.org/2023/02/07/covid-19-cases-and-deaths-by-age-jan-2020-feb-3-2023/ (accessed 15 July 2023).
[12] Jafar A, Dambul R, Dollah R, Sakke N, Mapa MT, *et al.* COVID-19 vaccine hesitancy in Malaysia: Exploring factors and identifying highly vulnerable groups. *PloS One* 2022, 17(7):e0270868.
[13] Vaithilingam S, Hwang LA, Nair M, Ng JWJ, Ahmed P, *et al*. COVID-19 vaccine hesitancy and its drivers: An empirical study of the vaccine hesitant group in Malaysia. *Plos One.* 2023, 18(3):e0282520.
[14] Reporters FMT. 35% of adults yet to receive booster shots, says Khairy. Available: https://www.freemalaysiatoday.com/category/nation/2022/03/16/35-of-adults-yet-to-receive-booster-shots-says-khairy/ (accessed 13 July 2023).
[15] Elnaem MH, Mohd Taufek NH, Ab Rahman NS, Mohd Nazar NI, Zin CS, *et al.* COVID-19 Vaccination Attitudes, Perceptions, and Side Effect Experiences in Malaysia: Do Age, Gender, and Vaccine Type Matter? *Vaccines* 2021, 9(10):1156.
[16] Wong LP, Alias H, Wong YL, Megat Hashim MMAA, Lin Y, *et al.* COVID-19 Anti-Vaccine Sentiments in Malaysia: Narratives of Comments from Facebook Post. *Vaccines* 2023, 11(4):834.
[17] Lee KW, Yap SF, Ong HT, Oo M, Swe KMM. COVID-19 Vaccine Booster Hesitancy in Malaysia: A Web-Based Cross-Sectional Study. *Vaccines* 2023, 11(3):638.
[18] Malaysia M of H & D of S. COVID-19 Vaccination | KKMNOW. Available: https://data.moh.gov.my/covid-vaccination (accessed 13 July 2023).
[19] Mohamed NA, Solehan HM, Rani MDM, Ithnin M, Arujanan M. Understanding COVID-19 vaccine hesitancy in Malaysia: Public perception, knowledge, and acceptance. *Plos One* 2023, 18(4):e0284973.
[20] Marzo RR, Ahmad A, Abid K, Khatiwada AP, Ahmed A, *et al.* Factors influencing the acceptability of COVID-19 vaccination: A cross-sectional study from Malaysia. *Vacunas (English Edition)* 2022, 23:33–40.
[21] Wong LP, Alias H, Siaw YL, Muslimin M, Lai LL, *et al.* Intention to receive a COVID-19 vaccine booster dose and associated factors in Malaysia. *Hum. Vaccines Immunother.* 2022, 18(5):2078634.

[22] Kennedy A, LaVail K, Nowak G, Basket M, Landry S. Confidence About Vaccines In The United States: Understanding Parents' Perceptions. *Health Aff*. 2011, 30(6):1151–1159.
[23] Yassine HM, Shah Z. How could artificial intelligence aid in the fight against coronavirus? *Expert Rev. Anti Infect. Ther.* 2020, 18(6):493–497.
[24] Wang L, Zhang Y, Wang D, Tong X, Liu T, *et al.* Artificial Intelligence for COVID-19: A Systematic Review. *Front. Med.* 2021, 8:1457.
[25] Bohloli H, Jamshidi HR, Ebraze A, Rabbani Khah F. Combining government, non-pharmaceutical interventions and vaccination in optimal control COVID-19. *Int. J. Healthc. Manag.* 2023, 16(1):61–69.
[26] Khan FM, Gupta R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *J. Saf. Sci. Resil.* 2020, 1(1):12–18.
[27] Our World in Data. Available: https://ourworldindata.org (accessed 15 July 2023).