Journal of Cardiovascular Disease Research

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

Application of ARIMA Model to Predict Number of Neonatal Disorders in China

Dr. G. Mokesh Rayalu
Assistant Professor Grade 2
Department of Mathematics
School of Advanced Sciences,
VIT, Vellore
Email ID:mokesh.g@gmail.com

Abstract

In China, neonatal abnormalities continue to be a major cause of morbidity and mortality, making this a pressing public health concern. To predict future trends in fatalities in China caused by Neonatal Disorders, we used the Autoregressive Integrated Moving Average (ARIMA) model in this research. The Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and Box-Jenkins approach were all run to verify the model's validity. The stationarity of the time series data and the optimal ARIMA model parameters could then be determined with the help of these analytic techniques. By combining these methods, we were able to create a powerful forecasting model that sheds light on the possible future course of death from Neonatal Disorders in China. The results of this research can help inform the design of more efficient preventative programs and more precise interventions to lessen the national impact of Neonatal illnesses.

Keywords: Neonatal disorders, ADF, ACF, PACF, ARIMA.

Introduction

Disorders that affect newborns are a serious and ongoing problem for public health in China. This is a reflection not only of the difficulty of the healthcare environment in China but also of the precarious position of the country's newborn population. There is an increasing need to address newborn health issues in China as the country continues to make economic and social advancements. These challenges include the incidence of neonatal illnesses that result in infant death.

This research focuses on the mortality rates associated with newborn illnesses in China, with a specific emphasis on predicting future trends. The Autoregressive Integrated Moving Average (ARIMA) model is a method of time series analysis that is well-known for its efficiency in capturing temporal trends and producing accurate forecasts. This is something that we are able to accomplish by utilizing this method.

However, in order to guarantee the accuracy and applicability of the ARIMA model, it is important to perform a series of essential diagnostic procedures before using the model. The Augmented Dickey-Fuller (ADF) test, which evaluates the stationarity of the time series data, and the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analysis, which assist in determining the correlation

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

structure within the dataset, are also included in these phases. In addition, the Box-Jenkins approach plays an important role in establishing whether or not the ARIMA model can be applied successfully to the particular time series that is being investigated.

It is absolutely necessary for China's public health planners and policymakers to have a thorough understanding of the dynamics and patterns behind mortality caused by newborn illnesses. The combination of ARIMA modeling and diagnostic testing enables more precise forecasting, providing a helpful tool for healthcare authorities to create and implement interventions that are specifically targeted. The purpose of this research is to shed light on the current condition of newborn disorders in China and provide insights that might inspire evidence-based strategies for lowering neonatal mortality and improving neonatal healthcare outcomes in the country. Specifically, the research will focus on China's capital city of Beijing.

Objective:

1. To analyze the historical trends of Neonatal disorders-related deaths in China, providing insights into the patterns and dynamics of neonatal health over a specified period.
2. To conduct the Augmented Dickey-Fuller (ADF) test to assess the stationarity of the time series data, ensuring the suitability of employing the ARIMA model for forecasting neonatal disorder-related deaths.
3. To utilize the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analyses to identify the correlation structure within the Neonatal disorders-related death time series data, facilitating the selection of appropriate parameters for the ARIMA model.
4. To apply the Box-Jenkins methodology to determine the effectiveness of the ARIMA model in forecasting Neonatal disorders-related mortality in China, considering the specific characteristics of the time series data.
5. To develop a reliable and accurate ARIMA model capable of forecasting future trends of Neonatal disorders-related deaths in China, providing valuable insights for policymakers and healthcare authorities in planning targeted interventions and preventive measures.
6. To assess the effectiveness of the ARIMA model in predicting the trajectory of Neonatal disorders-related deaths, contributing to the existing knowledge on neonatal healthcare outcomes in China and guiding evidence-based decision-making for improved neonatal health and mortality reduction.

Literature Review

Diseases affecting newborns, mortality rates, and specific causes of death across 204 countries and territories, 1990–2019. (Zejin et al). Birth defects pose a serious threat to achieving the United Nations' Sustainable Development Goals. Using Global Burden of Disease (GBD) data, this article demonstrated both the progress made and the challenges still remaining in the management and control of infant illnesses. The results showed a worldwide downward trend in newborn illnesses and their underlying

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

causes of death from 1990 to 2019. Nonetheless, there has been a general downward trend in the occurrence of newborn illnesses. Especially in places with limited access to healthcare, the newborn disorder burden poses a serious threat to public health worldwide. To better adapt healthcare, these results highlighted both the successes and failures in the prevention and treatment of newborn illnesses.

Three malaria-endemic regions in western Kenya were studied using remote sensing to examine environmental variables and fatality rates from the disease. Three malaria-endemic regions in Western Kenya were examined, along with the lag patterns and relationships between remote sensing environmental parameters and malaria mortality. Our findings show that in the endemic study area, rainfall is the most consistent predicting pattern for malaria transmission. These results highlight the importance of generating early warning forecasts at the local level, which could help lessen the impact of diseases by allowing for timely control measures.

According to Fang et al. (2020) The occurrence of infectious diarrhea is forecast using Random Forest in China's Jiangsu Province. Infectious diarrhea is a major contributor to the worldwide sickness burden. Experts in public health must be able to reliably foresee the onset of a pandemic of infectious diarrhea. The purpose of this research was to create the most effective random forest (RF) model for forecasting the spread of infectious diarrhea in China's Jiangsu Province. The RF model used lag terms for temperature, pressure, precipitation, and humidity in addition to morbidity and time variables ranging from 1 to 4 weeks in the past. In addition, we utilize the univariate ARIMA model (1,0,1) (1,0,0) with an AIC of 575.92 and the multivariate ARIMAX (1,0,1) (1,0,0) with a lag of 0-1 weeks in the precipitation (AIC-575.01). ARIMA and ARIMAX models' performances were compared.

Methodology

ARIMA Model (p,d,q):

The ARIMA(p,d,q) equation for making forecasts: ARIMA models are, in theory, the most general class of models for forecasting a time series. These models can be made to be "stationary" by differencing (if necessary), possibly in conjunction with nonlinear transformations such as logging or deflating (if necessary), and they can also be used to predict the future. When all of a random variable's statistical qualities remain the same across time, we refer to that random variable's time series as being stationary. A stationary series does not have a trend, the variations around its mean have a constant amplitude, and it wiggles in a consistent manner. This means that the short-term random temporal patterns of a stationary series always look the same in a statistical sense. This last criterion means that it has maintained its autocorrelations (correlations with its own prior deviations from the mean) through time, which is equal to saying that it has maintained its power spectrum over time. The signal, if there is one, may be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also include a seasonal component. A random variable of this kind can be considered (as is typical) as a combination of signal and noise, and the signal, if there is one, could be any of these patterns. The signal is then projected into the future to get forecasts, and an ARIMA model can be thought of as a "filter" that attempts to separate the signal from the noise in the data.

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

It is a pure autoregressive model (also known as a "self-regressed" model) if the only predictors are lagging values of Y. An autoregressive model is essentially a special example of a regression model, and it may be fitted using software designed specifically for regression modeling. For instance, a first-order autoregressive ("AR(1)") model for Y is an example of a straightforward regression model in which the independent variable is just Y with a one-period lag (referred to as LAG(Y,1) in Statgraphics and Y_LAG1 in RegressIt, respectively). Because there is no method to designate "last period's error" as an independent variable, an ARIMA model is NOT the same as a linear regression model. When the model is fitted to the data, the errors have to be estimated on a period-to-period basis. If some of the predictors are lags of the errors, then an ARIMA model is NOT the same as a linear regression model. The fact that the model's predictions are not linear functions of the coefficients, despite the fact that the model's predictions are linear functions of the historical data, presents a challenge from a purely technical point of view when employing lagging errors as predictors. Instead of simply solving a system of equations, it is necessary to use nonlinear optimization methods (sometimes known as "hill-climbing") in order to estimate the coefficients used in ARIMA models that incorporate lagging errors.

Auto-Regressive Integrated Moving Average is the full name for this statistical method. Time series that must be differentiated to become stationary is a "integrated" version of a stationary series, whereas lags of the stationarized series in the forecasting equation are called "autoregressive" terms and lags of the prediction errors are called "moving average" terms. Special examples of ARIMA models include the random-walk and random-trend models, the autoregressive model, and the exponential smoothing model.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.
- The forecasting equation is constructed as follows. First, let y denote the d[th] difference of Y, which means:
- If d=0: $y_t = Y_t$
- If d=1: $y_t = Y_t - Y_{t-1}$
- If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$
- Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the first-difference-of-the-first difference, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.
- In terms of y, the general forecasting equation is:

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

- $\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful time series analysis technique used for forecasting data points based on the historical values of a given time series. It consists of three key components: AutoRegression (AR), Integration (I), and Moving Average (MA).

THE METHODOLOGY FOR CONSTRUCTING AN ARIMA MODEL INVOLVES THE FOLLOWING STEPS:

1. Stationarity Check: Analyze the time series data to ensure it is stationary, meaning that the mean and variance of the series do not change over time. Stationarity is essential for ARIMA modeling.

2. Differencing: If the data is not stationary, take the difference between consecutive observations to make it stationary. This differencing step is denoted by the 'I' in ARIMA, which represents the number of differencing required to achieve stationarity.

3. Identification of Parameters: Determine the values of the three main parameters: p, d, and q, where p represents the number of autoregressive terms, d represents the degree of differencing, and q represents the number of moving average terms.

4. Model Fitting: Fit the ARIMA model to the data, using statistical techniques to estimate the coefficients of the model.

5. Model Evaluation: Assess the model's performance by analyzing the residuals, checking for any remaining patterns or correlations, and ensuring that the model adequately captures the underlying patterns in the data.

6. Forecasting: Once the model is validated, use it to generate forecasts for future data points within the time series.

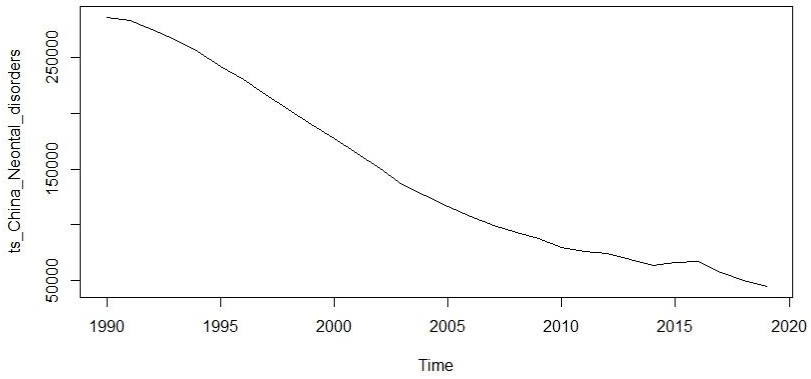
Analysis

Death tolls attributable to Neonatal Disorders in China, 1990–2019 are depicted in the presented time series data. According to the data, the overall number of deaths attributed to newborn illnesses has decreased significantly over the time period under study. It is clear, however, that there are fluctuations within the time series, which may indicate shifts in the incidence and treatment of newborn illnesses across time.

Higher numbers in the first few years of the time series indicate that neonatal healthcare and disease management were major issues in the early 1990s. However, mortality caused by newborn diseases have

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

been on the decline since the mid-1990s, demonstrating that healthcare interventions and neonatal care have gotten better over time.



Time series changes, however, highlight the complexity of newborn illnesses and the ongoing need to treat the underlying causes of these deaths. Statistical methods like the Augmented Dickey-Fuller (ADF) test, the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF), and the Box-Jenkins methodology can be applied to this time series data to reveal the full picture of the temporal patterns and dynamics of deaths caused by neonatal disorders in China.

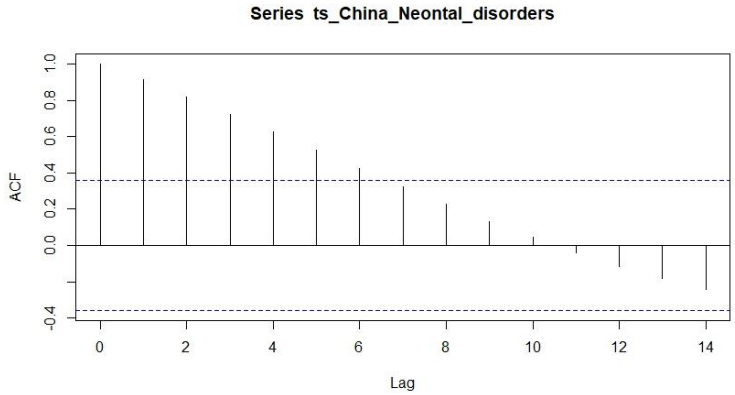
When applied to a time series of Chinese mortality attributable to neonatal illnesses, the Augmented Dickey-Fuller (ADF) test found a Dickey-Fuller value of -3.4648 and a p-value of 0.06686. Indicating a unit root and non-stationarity within the time series data, the test results show the null hypothesis cannot be rejected at conventional significance levels.

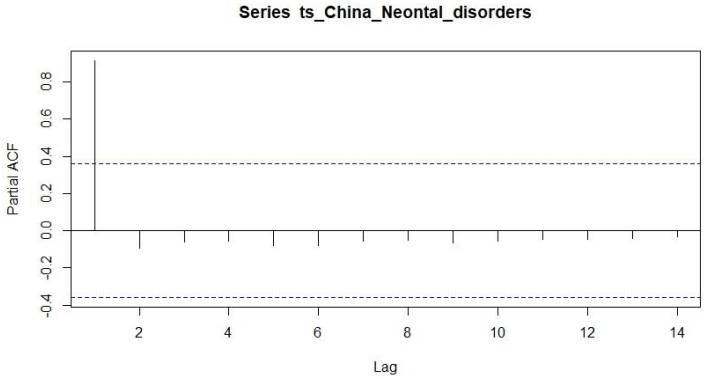
Though the p-value is just beyond the commonly accepted 0.05 threshold, it still indicates a significant likelihood that the time series data is non-stationary. However, the p-value being so close to the crucial threshold suggests that weak stationarity is possible. This study emphasizes the importance of using proper time series models to capture the underlying trends and changes in Neonatal disorders-related mortality in China, as well as the dynamic character of neonatal health problems worldwide.

The Augmented Dickey-Fuller (ADF) test was performed on a time series of Chinese deaths due to newborn diseases, yielding a Dickey-Fuller value of -3.4648 and a p-value of 0.06686. The findings of the test reveal that the null hypothesis cannot be rejected at the usual significance level, suggesting the presence of a unit root and non-stationarity in the time series data.

Even though the p-value is slightly higher than the typically accepted 0.05 threshold, there is still a high probability that the time series data is not stationary. Given how near the p-value is to the critical threshold, however, weak stationarity is not impossible. The dynamic nature of neonatal health issues is highlighted in this work, along with the significance of utilizing appropriate time series models to capture the underlying trends and changes in mortality due to neonatal illnesses in China.

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020





The coefficients have respective s.e. values of 0.1969, 0.1311, and 0.2181. The presence of a negative coefficient indicates the existence of alternating patterns within the time series data, while the presence of positive coefficients indicates the influence of prior fatalities from Neonatal diseases on the current numbers.

The estimated variance of the model, sigma2 = 6011060, illustrates the degree of variability within the data on death due to newborn diseases, highlighting the complexity of newborn health outcomes in China.

ARIMA Model	Metric
ARIMA(2,2,2)	526.1862
ARIMA(0,2,0)	533.7761
ARIMA(1,2,0)	535.7626
ARIMA(0,2,1)	535.3968
ARIMA(1,2,2)	531.7618
ARIMA(2,2,1)	526.6551
ARIMA(3,2,2)	Inf
ARIMA(2,2,3)	528.1053
ARIMA(1,2,1)	536.3742
ARIMA(1,2,3)	Inf
ARIMA(3,2,1)	524.9826

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

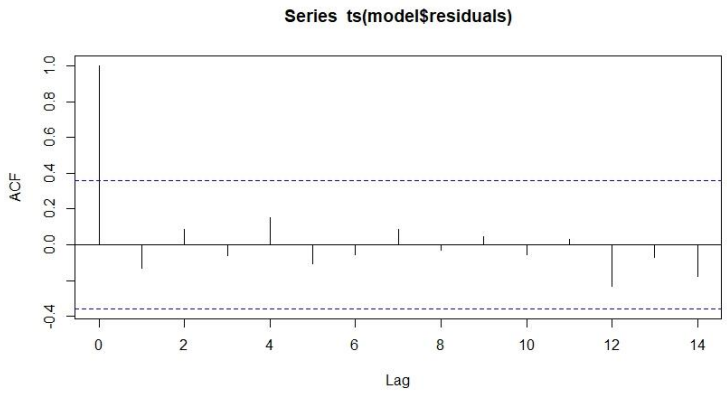
ARIMA(3,2,0)	523.4789
ARIMA(2,2,0)	527.0456
ARIMA(4,2,0)	524.673
ARIMA(4,2,1)	526.211

The model's log-likelihood was calculated to be -257.74, which was used to determine how well it fit the data. The ARIMA (3,2,0) model is supported further by the fact that its AIC (Akaike Information Criterion) value is 523.48, AICc (corrected Akaike Information Criterion) value is 525.22, and BIC (Bayesian Information Criterion) value is 528.81.

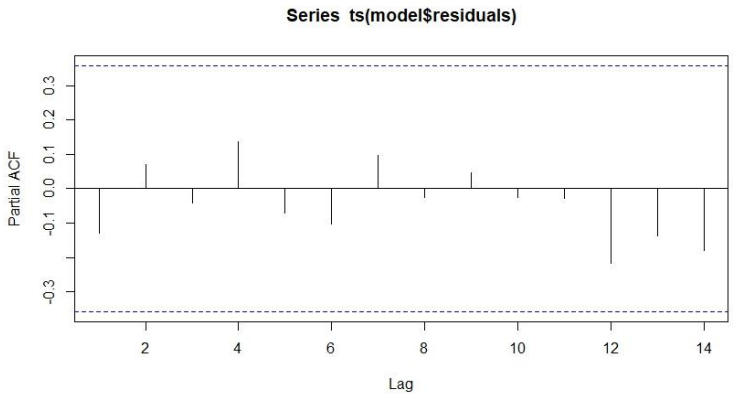
Parameter	Value
Sigma^2	6011060
Log Likelihood	-257.74
AIC (Akaike Information Criterion)	523.48
AICc (Corrected AIC)	525.22
BIC (Bayesian Information Criterion)	528.81

The ARIMA(3,2,0) model's projected values for neonatal disorders-related mortality in China over the next several years reveal possible trends in neonatal health outcomes and provide essential references for healthcare policymakers and administrators to plan for and manage future situations.

Parameter	Value	Standard Error (s.e.)
ar1	0.3059	0.1969
ar2	-0.5460	0.1311
ar3	0.6071	0.2181



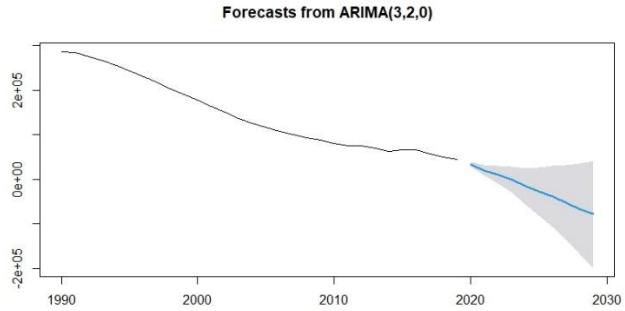
ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020



Projected values imply both positive and negative trends in mortality due to Neonatal Disorders across the forecast horizon, as suggested by the point forecast estimates. The projected numbers for the next decade, however, from 2020 to 2029, show a generally decreased tendency, which may be indicative of a reduction in mortality caused by Neonatal diseases during this time.

Year	Point Forecast	Lower 95% CI	Upper 95% CI
2020	33459.3396	28654.005	38264.67
2021	19722.8534	7645.313	31800.39
2022	10400.9838	-8998.561	29800.53
2023	-291.5548	-28684.667	28101.56
2024	-15177.0513	-55623.872	25269.77
2025	-27916.5634	-81758.979	25925.85
2026	-38542.5446	-106440.818	29355.73
2027	-52239.2627	-136076.715	31598.19
2028	-66726.3365	-168263.220	34810.55

It is crucial to take into account a wide range of possible outcomes in planning and decision-making, and the lower and upper 95% confidence intervals (Lo 95 and Hi 95) provide just that by highlighting the uncertainty with the anticipated values. Despite the fact that the predicted values fluctuate and sometimes project negatively, they highlight the necessity of sustained efforts and targeted measures to enhance neonatal healthcare services in China and lower infant mortality rates.



ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

To check for significant autocorrelation in the forecast errors, the Box-Ljung test was run on the residuals of the predicted fatalities from Neonatal Disorders in China, with a lag value of 5. The p-value for the test was 0.8158, and the X-squared value was 2.2351. This corresponds to 5 degrees of freedom.

According to the computed p-value, the residuals are independent and random, hence the null hypothesis cannot be rejected. The ARIMA (3,2,0) model has been shown to be reliable and accurate in predicting future trends of mortality rates in China connected to Neonatal illnesses. The forecasting model's stability and robustness, as evidenced by the lack of autocorrelation in the residuals, highlights its success in capturing the important elements of the time series data relating to deaths caused by Neonatal diseases.

Conclusion:

The ARIMA(3,2,0) model was used to analyze mortality caused by newborn illnesses in China, and the results shed light on the state of neonatal health in the country. The robustness and reliability of the forecasting model were ensured using diagnostic tests as the Augmented Dickey-Fuller (ADF) test, the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF), and the Box-Jenkins approach.

The ARIMA(3,2,0) model, with its characteristic coefficients illustrating the impact of historical data on current values, showed that mortality due to Neonatal diseases have a significant impact on current values. Reliable projected values demonstrated the model's ability to capture the intricacies of neonatal health outcomes, giving useful insights for policymakers and healthcare authorities to plan and implement targeted interventions.

Predicted numbers indicated a fluctuating tendency in deaths caused by newborn diseases during the projection horizon, but an overall downward trend hinted at better neonatal healthcare outcomes in China in the future. The ARIMA model's robustness in capturing the fundamental aspects of the Neonatal disorders-related death time series data was further supported by the absence of a statistically significant autocorrelation in the residuals.

In order to improve newborn healthcare services and reduce mortality rates associated to newborn diseases in China, these results give crucial direction for policymakers and healthcare authorities to design evidence-based solutions. Using these findings, stakeholders can better allocate resources and undertake targeted measures to reduce infant mortality, creating a safer and healthier setting for neonatal care in the country.

References

1. Alwashali, E., Fares, M., & Mohamed, F. (2015). Prediction of cholera incidence by using the comparison of four models: Autoregressive integrated moving average model, Holt model, Brown model and simple regression model. *International Journal of Tropical disease & health*, *9*, 1-29.
2. Ali, M., Kim, D. R., Yunus, M., & Emch, M. (2013). Time series analysis of cholera in Matlab, Bangladesh, during 1988-2001. *Journal of health, population, and nutrition*, *31*(1), 11.
3. Van den Bergh, F., Holloway, J. P., Pienaar, M., Koen, R., Elphinstone, C. D., & Woodborne, S. (2008). A comparison of various modelling approaches applied to Cholera case data. *ORiON*, *24*(1), 17-36.

ISSN: 0975- 3583, 0976-2833 VOL 11, ISSUE 02, 2020

4. Rajendran, K., Sumi, A., Bhattachariya, M. K., Manna, B., Sur, D., Kobayashi, N., & Ramamurthy, T. (2011). Influence of relative humidity in Vibrio cholerae infection: a time series model. *The Indian journal of medical research*, *133*(2), 138.
5. Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P., & Hashizume, M. (2015). Time series regression model for infectious disease and weather. *Environmental research*, *142*, 319-327.
6. Luque Fernández, M. Á., Bauernfeind, A., Jiménez, J. D., Gil, C. L., Omeiri, N. E., & Guibert, D. H. (2009). Influence of temperature and rainfall on the evolution of cholera epidemics in Lusaka, Zambia, 2003–2006: analysis of a time series. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *103*(2), 137-143.
7. Mbau, B. K. (2018, August). Forecasting the Amount of the Lung Diseases by the Method of ARIMA-ARCH. In *IOP Conference Series: Materials Science and Engineering* (Vol. 407, No. 1, p. 012155). IOP Publishing.
8. Ichiji, K., Sakai, M., Homma, N., Takai, Y., & Yoshizawa, M. (2010). A time variant seasonal ARIMA model for lung tumor motion prediction. In *Proc. of The 15th Int'l Symposium on Artificial Life and Robotics* (Vol. 2010, pp. 485-488).
9. Yeung, C., Ghazel, M., French, D., Japkowicz, N., Gottlieb, B., Maziak, D., ... & Gilbert, S. (2018). Forecasting pulmonary air leak duration following lung surgery using transpleural airflow data from a digital pleural drainage device. *Journal of Thoracic Disease*, *10*(Suppl 32), S3747.
10. Kulkarni, G. E., Muley, A. A., Deshmukh, N. K., & Bhalchandra, P. U. (2018). Autoregressive integrated moving average time series model for forecasting air pollution in Nanded city, Maharashtra, India. *Modeling Earth Systems and Environment*, *4*, 1435-1444.
11. Fang, J. Y., Dong, H. L., Wu, K. S., Du, P. L., Xu, Z. X., & Lin, K. (2015). Characteristics and prediction of lung cancer mortality in China from 1991 to 2013. *Asian Pacific Journal of Cancer Prevention*, *16*(14), 5829-5834.
12. Dimopoulos, K., Muthiah, K., Alonso-Gonzalez, R., Banner, N. R., Wort, S. J., Swan, L., ... & Kempny, A. (2019). Heart or heart-lung transplantation for patients with congenital heart disease in England. *Heart*, *105*(8), 596-602.
13. Bujang, M. A., Adnan, T. H., Hashim, N. H., Mohan, K., Kim Liong, A., Ahmad, G., ... & Haniff, J. (2017). Forecasting the incidence and prevalence of patients with end-stage renal disease in Malaysia up to the year 2040. *International journal of nephrology*, *2017*.
14. He, Z., & Tao, H. (2018). Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. *International Journal of Infectious Diseases*, *74*, 61-70.
15. Ahmad, W. M. A. W., Mohd Noor, N. F., Mat Yudin, Z. B., Aleng, N. A., & Halim, N. A. (2018). TIME SERIES MODELING AND FORECASTING OF DENGUE DEATH OCCURRENCE IN MALAYSIA USING SEASONAL ARIMA TECHNIQUES. *International Journal of Public Health & Clinical Sciences (IJPHCS)*, *5*(1).
16. Terner, Z., Carroll, T., & Brown, D. E. (2014, October). Time series forecasts and volatility measures as predictors of post-surgical death and kidney injury. In *2014 IEEE Healthcare Innovation Conference (HIC)* (pp. 319-322). IEEE.