

Medical Full Length Article

Modeling and prediction of hypertension in Komfo Anokye Teaching Hospital (KATH), Ghana

Agyei Helena Lartey¹, Jianming Wang¹, Philip Lartey³, James Agyei², Alex Agyei², Janet Sintim Aboagye²

¹Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, PR China

²Komfo Anokye Teaching Hospital, Laboratory Services Directorate Chemical Pathology Unit

³Eastern Regional Hospital, Herbal Medicine Department, Koforidua, Ghana

Correspondence: Jianming Wang. Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, 101 Longmian Ave. Nanjing, 211166, China. Tel: +86-25-86868438. Email:

jmwang@njmu.edu.cn

Accepted 18th June, 2020

To explore the temporal trends of hypertension in a Ghana population and to predict future values, which will, in turn, help control and reduce the risk of hypertension-related health events. We enrolled 108,100 cases with essential hypertension from January 2015 to December 2019 at the Komfo Anokye Teaching Hospital (KATH), Ghana. The Box-Jenkins Autoregressive Integrated Moving Average model (ARIMA) was used to identify trends and forecast data from a specified time series. The root mean square error (RMSE), Q-statistic, residual variance (RV), and Akaike's information criteria (AIC) were used to assess the performance of the model. The most optimal model was ARIMA(1, 1, 0) with RV(7061), RMSE(82.6155), AIC(693.48), Q-value(19.187), parameter(-0.4034) and constant(188.6501). The best fitting model was $Y_t = (1-0.4034)Y_{t-1} - 0.4034Y_{t-2} + 1801.6670$. The model estimated an increase in hypertension cases for the next period, which was a critical input in managerial and administrative decision making. The forecast was accurate enough to allow for better planning and control than could be accomplished without the forecast.

Keywords: Hypertension, Forecast, ARIMA, RMSE, RV, AIC

INTRODUCTION

Forecasting future health events facilitates preventive medicine and health care intervention strategies by pre-informing health service providers to take proper measures to control health conditions. In the recent global burden of disease study, hypertension affects more than one billion adults in the world, and Ghana is not an exception. Hypertension significantly increases the risks of damage to the heart, brain, kidney, and other organs (Mendis et al. 2010; Mendis et al. 2007). In 2015, 1 in 4 men and 1 in 5 women had hypertension. Also, fewer than 1 in 5 people have hypertension under control. Hypertension is a major cause of premature death worldwide, and as a result, one of the global targets for non-communicable diseases is to reduce the prevalence of hypertension by 25% by 2025 (baseline 2010) (WHO 2017; WHO 2019).

Hypertension contributes about 55% of the global mortality caused by cardiovascular diseases and 7% of all disability-adjusted life years. In the developed world, more than 80% of people with hypertension are aware of their condition and receiving treatment (Wilkins et al. 2010). However, the health systems in most developing countries fail to early detect and manage hypertension effectively (Basu and Millett 2013; Mohsen Ibrahim 2018). According to the

Ghana Health Service (GHS), one of the main barriers to effective control of hypertension is that people are not aware of their blood pressure, which makes it the second leading cause of outpatient morbidity in adults older than 45 years (GHS 2019; Mohsen Ibrahim 2018). In 2015, a nationwide survey conducted in China revealed that 244.5 million adults (23%) had hypertension, and another 435 million (41.3%) had prehypertension. Of those with hypertension, only 47% were aware of their condition, and fewer than 18% of such individuals effectively controlled their blood pressure (Li and Wang 2012; Wang et al. 2019).

Hypertension not only causes premature death; it may also add to household costs (Basu and Millett 2013; He et al. 2009; Le et al. 2012). In a study conducted in rural China, for example, it was estimated that 4.1% of households suffered impoverishment as a result of hypertension (Le et al. 2012). The study on the prevalence and early diagnosis of hypertension in adults is an essential strategy for the public control and prevention of cardiovascular diseases (Anker et al. 2020; Gu et al. 2005). It is, therefore, vital to explore the pattern of hypertension among Ghanaians to predict future values, which will, in turn, help more effective measures to be taken to improve health.

METHODS

Study subjects

Data were extracted from Komfo Anokye Teaching Hospital (KATH), Ghana, from January 2015 to December 2019. KATH is located at the center of Kumasi, the regional capital of the Ashanti Region. Kumasi has an estimated population of 3.2 million. It is about 10.2% for the total landmass of Ghana of size 24,390sq km. KATH is a teaching hospital and receives referrals from both private and public health sectors within the metropolis, all the 19 districts within the region and beyond. It was therefore considered appropriate to study the temporal pattern of high blood pressure among patients with a view to effectively integrate the forecasting model into the existing disease control program in reducing the occurrence of its associated health events. Systolic and diastolic blood pressure (BP) for 108,100 patients with hypertension were extracted from the daily hospital attendance at the outpatient department. The study period was divided into 61 months. Data used in this study were anonymized and irreversibly de-identified to protect patients, health care professionals, and hospital privacy.

Blood pressure measurement

Categories of hypertension were also estimated according to the 2017 American College of Cardiology/American Heart Association High Blood Pressure Guideline (Whelton et al. 2018). We defined hypertension as blood pressure at or above 140/90 mmHg. Blood pressure was measured using a standard mercury sphygmomanometer, and values were approximated to the nearest number (mmHg).

STATISTICAL ANALYSIS

Modeling

Time series is a time-dependent sequence such as Y_1, Y_2, \dots, Y_n , where n denotes time steps which can be deterministic or stochastic, has expanded in recent years. The Box-Jenkins approach to forecasting was first described by George Box and Gwilym Jenkins based on a particular class of linear stochastic models (Box 1976). In the current study, time series analysis was performed using the Box-Jenkins method. Autoregressive Integrated Moving Average model (ARIMA) was applied to identify trends and forecast data from a specified time series. The ARIMA model was identified through the autocorrelation and partial autocorrelation functions from the ARIMA family (i.e., $\alpha(B)\nabla^d Y_t = \theta(B)\epsilon_t$), which was used to represent the time series. Time-series modeling with the Box-Jenkins approach has to be stationary without trend, having a constant mean and variance over time, in order to be used to predict future values.

The first-order differencing was required to remove the trend in mean to obtain stationarity in the data ($Y_t = \text{hypertension cases}$) as $Y_t = X_{t+1} - X_t = \Delta X_{t+1}$ and transforming $Y_t = U_t$ where $U_t = \log X_t$ ensured stationarity in the variance. An autoregressive model of order p denoted by $AR(p)$ was represented by $Y_t = \sum_{i=1}^p \alpha_i Y_{t-i}$ or $Y_t = \sum_{k=1}^p \alpha_k Y_{t+k} + \mu + \epsilon_t$, where μ was the mean of the time series data, and ϵ_t was the white noise. Also, the backward-shift operator, B , operated on the series to move it back one time unit as $B_k Y_t = Y_{t-k}$. Partial autocorrelation function determined the order of an $AR(p)$ process and measured the degree of association between Y_t and Y_{t+k} at various lags k was defined by $\Phi_{kk} = |P_k| / |P_K|$, where P_k was $k \times k$ autocorrelation matrix. For AR models, the ACF dampened exponentially, and the

PACF was used to identify the order (p) of the AR model. The number of significant spikes on the PACF determined the AR model order (p).

Moving average model MA(q) provided predictions of Y based on a linear combination of past forecast errors. The q was given by $Y_t = \sum_{k=1}^q \theta_k e_{t-k} + \mu + e_t$. For MA models, the PACF dampened exponentially, and the ACF plot was used to identify the order of the MA process. The number of significant spikes on the ACF determined the MA model order (q). ARIMA (p,d,q), was derived from ARMA (p, q) when the non-stationary time series was differenced to remove the variation, to obtain an integrated time series. Rationally, all AR(p) and MA(q) models could be represented as ARIMA(p,d,q), where p was the order of the AR part, d the degree for differencing and q the order of the MA part. The general ARIMA model was of the form $W_t = \sum_{k=1}^p \alpha_k W_{t-k} + \sum_{l=1}^q \theta_l e_{t-l} + u + e_t$. The parameters (p, d, q) were estimated and the accuracy of the model used to forecast the values of the test data set was determined by assessing:

Q-statistic

$$Q = n(n+2) \sum_{k=1}^K \frac{r_k}{n-k} (\rho) \approx \text{chi-square}$$

It is approximately distributed as a chi-square with k-p-q degrees of freedom where n is the length of the time series, k is the first k autocorrelations being checked, p is the order of the AR process, and q is the order of the MA process, and r is the estimated autocorrelations coefficients of the residual term. If the calculated value of Q is higher than chi-square for k-p-q degrees of freedom, then the model is considered inadequate and adequate if Q is less than for k-p-q degrees of freedom.

Parsimony

Parsimony is concerned about the number of parameters in a time series model. The fewer the number of parameters to estimate, the simpler and more parsimonious the time series model is for adequate representations. When choosing from stationary competing models, ARIMA(0,0,2) and ARIMA(1,0,0), ARIMA(1,0,0) is selected over ARIMA(0,0,2).

Residual variance (RV)

Residual is the difference between the observation and the value estimated from the regression model. The source of the analysis may lead to variance unexplained by the regression that is error variance. The smaller residual variance e_t is selected for Q-ststistic value when modeling.

Akaike's information criteria (AIC)

The AIC uses the maximum likelihood method to estimate the range of ARMA models as $AIC(p, q) = \ln(\sigma^2) + \frac{2}{n}$ + constant, where n is the sample size or the number of observations, σ^2 is the maximum likelihood estimate, and the residual variance, $r = p+q+1$ denotes the number of parameters estimated in the model. Among stationary competing models, the one with the smaller AIC value will be selected. Augmented Dickey-Fuller (ADF) unit root test was used to measure the stationarity. The PACF and ACF coefficient of random data with $N \geq 40$ were between -0.316 and 0.316; hence any value outside this interval was considered to be significantly different from zero. The smaller residual variance, e_t was selected for the Q-value when modeling and the Q-statistic less than chi-square for k-p-q degrees of freedom was considered adequate. A more parsimonious model (fewer parameters) with smaller AIC was selected over stationary competing models. Statistical tests were set at a significance level of 0.05. All data analyses were conducted using R programming language.

RESULTS

As shown in Figure 1, the annual prevalent hypertension varied between years, with the mean of 1802 cases. The highest percentage deviation was recorded in 2019 whereas the least was recorded in 2016.

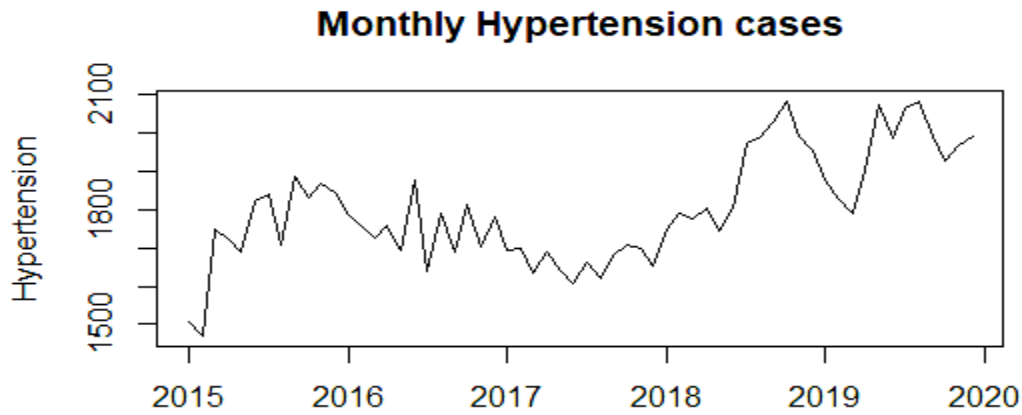


Figure 1. Raw plots showing Monthly cases of Hypertension over time

The time series was decomposed into $Y_t = S_t + T_t + E_t$. (1) Y_t , observed data-the actual data plot, (2) T_t , trend-the overall upward or downward movement of the data points which was differenced to achieve stationarity, (3) S_t , seasonal-monthly pattern of the data points, (4) E_t , random-unexplained part of the data. The P -value of 0.5351 from the ADF test indicated that the log returns for the series was non-stationary and required first differencing (de-trending) to make it stationary (Figure 2).

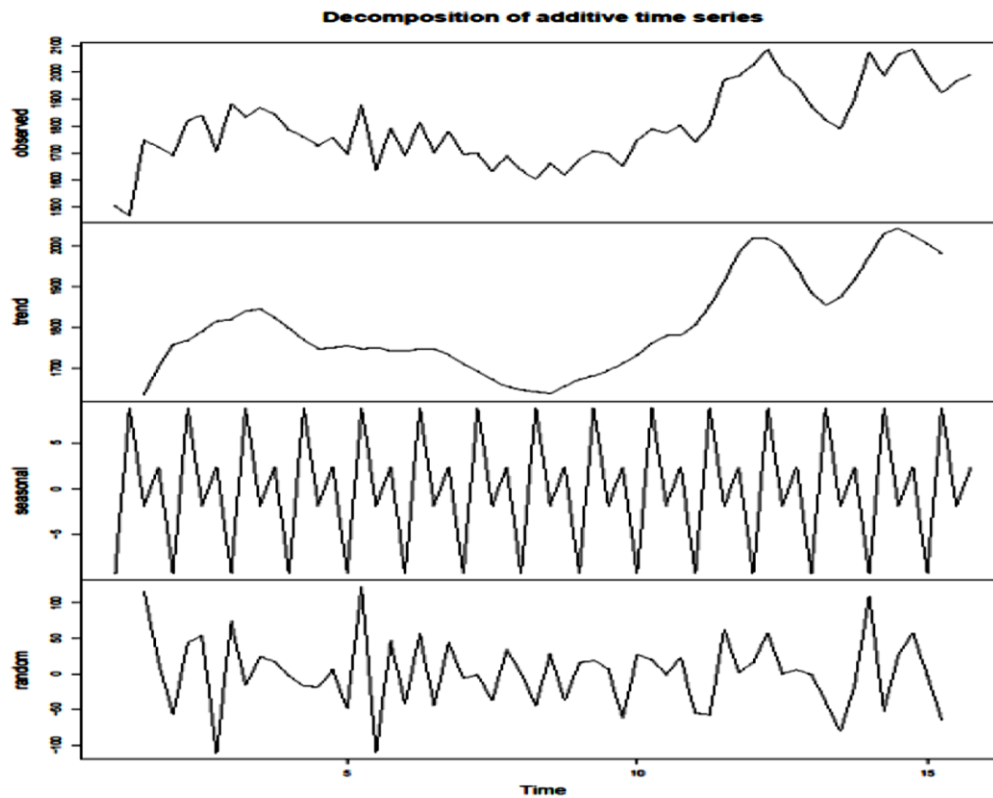


Figure 2: Graph showing decomposition of additive time series

The ACF dampened exponentially, and the PACF plot was used to identify the order (p) of the AR model. As shown in Figure 3, AR order =1 and MA order =0.

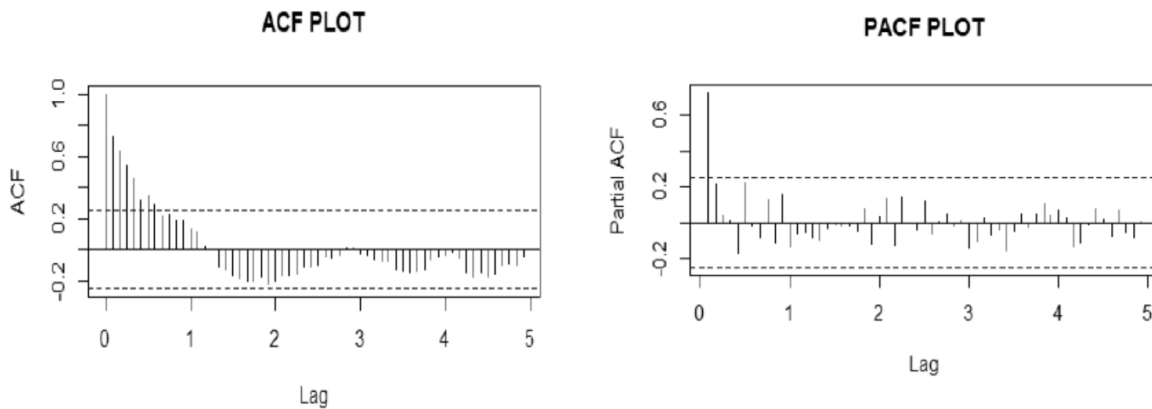


Figure 3. ACF and PACF of Hypertension cases

There was a first-order differencing; thus the ARIMA (1,1,0) was used to forecast the next data points. The ARIMA(1,0,0) had a better AIC (694.29), RV(6801), the smaller Q-value (17.190) and RMSE (82.47013) than the ARIMA (1,1,0) but if taken to forecast, there would be a greater variation from the actual values. This was because its AIC was greater than that of ARIMA (1, 1, 0), but a model with a smaller AIC is preferred to that of a greater one. Hence ARIMA (1, 1, 0) model with a smaller AIC is selected over ARIMA (1, 0, 0) (Table 1).

Table 1. Modeled data on collected hypertension cases

Parameter	ARIMA Model		
	(1,1,0)	(1,0,0)	(0,0,1)
Residual Variance(RV)	7061	6801	0.004024
Akaike's Information Criteria(AIC)	693.48	694.29	-154.36
Q-value	19.187	17.190	23.580
Stationarity	Yes	Yes	Yes
Parameter	-0.4034	-0.4135	0.4982
Root Mean Square Error(RMSE)	83.31242	82.47013	0.0634382
ACF1 Training set	-0.007332488	0.007911908	0.2718276

Source: Output from R programming Language

From the coefficients obtained from ARIMA (1,1,0), the return equation was $Y_t = (1 - 0.4034)Y_{t-1} - 0.4034Y_{t-2} + 1801.6670$. The standard errors for the coefficients were within the acceptable limits -0.316 and 0.316 . The AIC score, 693.48 was a good indicator of the ARIMA model accuracy. The ACF plot of the residuals had its autocorrelations below the threshold limit with $\chi^2 = 19.187$, $df = 20$, and $P\text{-value} = 0.5097$ suggested a good ARIMA model and satisfied the variability and normality conditions (Figure 4).

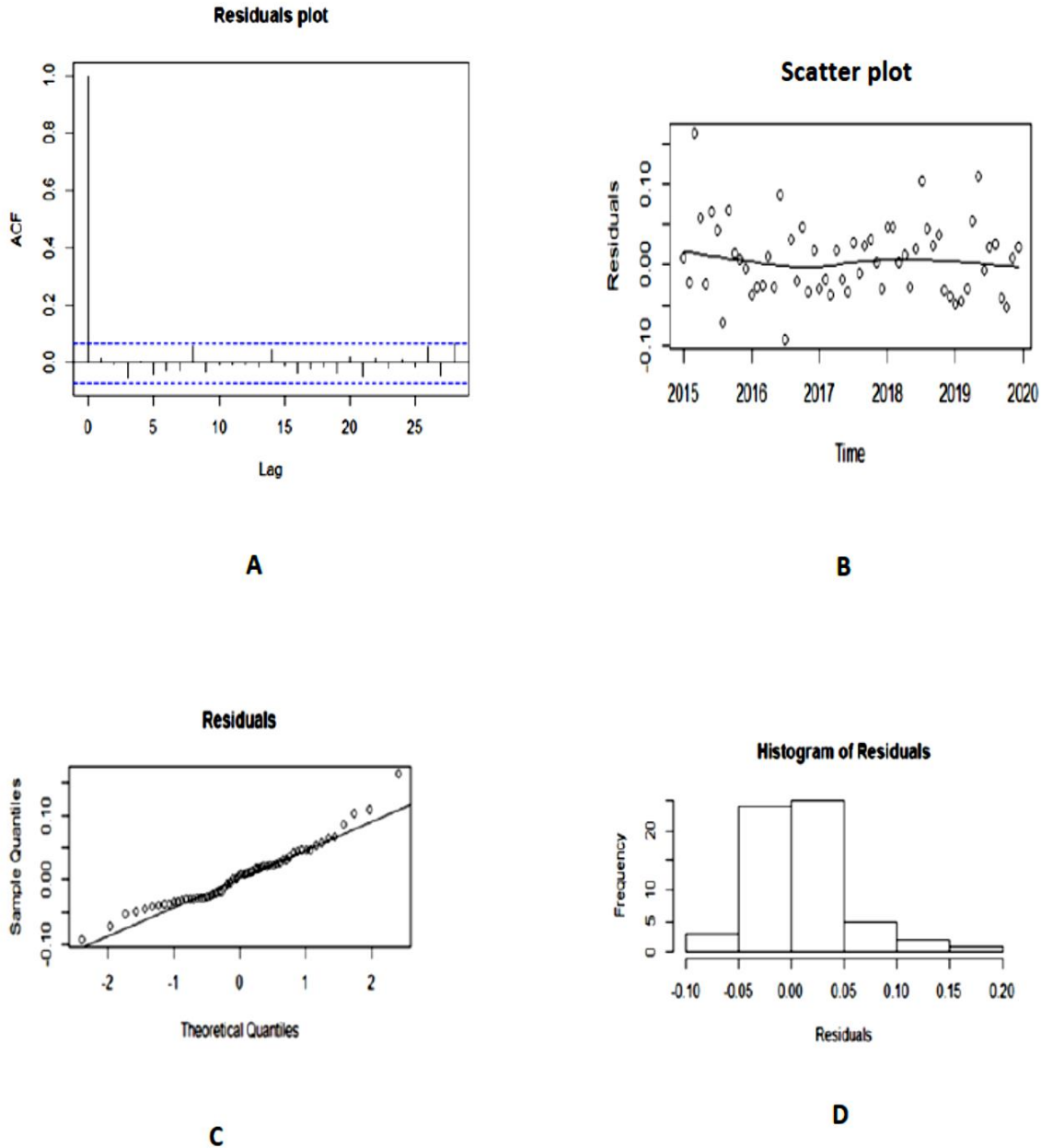


Figure 4. Multiple plots (A, B, C, D) showing residuals of an ARIMA (1, 1, 0)

The forecasted point return was -0.007332488 . The forecasted return fitted the actual returns, and forecasts from the model were adequate.

$$Y_t = (1 - 0.4034) Y_{t-1} - 0.4034 Y_{t-2} + 1801.6670$$

$$Y_{68} = (1 - 0.4034) Y_{67} - 0.4034 Y_{66} + 1801.6670$$

$$Y_{68} = (1 - 0.4034)(2218.41106) - 0.4034(2212.155381) + 1801.6670 = 2232.81907 \approx 2233 \text{ (Figure 5).}$$

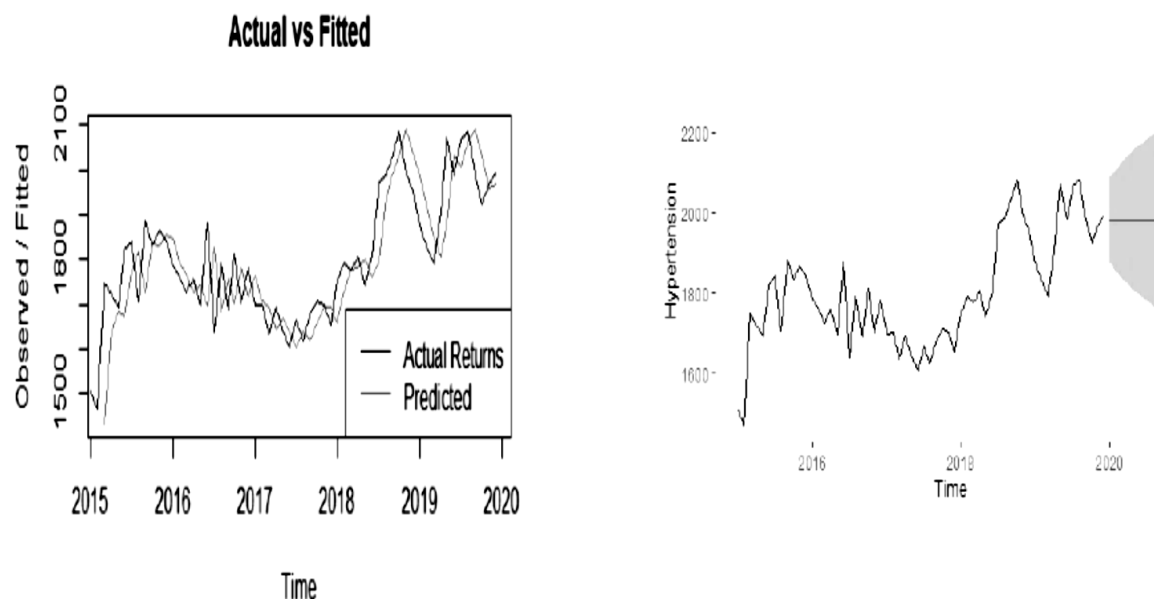


Figure 5. Forecasts of ARIMA (1, 1, 0) model

DISCUSSION

We performed a time-series study to explore the temporal trends of hypertension in a Ghana population and predicted future prevalence to help prevent, treat, control, and reduce the risk of hypertension-related health events. The hypertension cases will not only continue to occur but may even increase in the near future if timely effective and efficient intervention measures are not put in place. The ARIMA model used in this study presents the best feasible alternative for predicting hypertension cases per month in the future based on past observed hypertension cases over the years.

Different time-series models have been used to predict the trend of infectious diseases. A study in Zambia, using the Box-Jenkins modeling procedure to determine an ARIMA model of Malaria cases from 2009 to 2013 for age 1 to under 5 years reported that malaria cases were more likely to occur and increase in the near future if not controlled on time (Stanley Jere and Edwin Moyo 2016). In a surveillance time series data for primary, secondary, tertiary, congenital and latent syphilis in mainland China from 2005 to 2012 using Univariate ARIMA model and an autoregressive integrated moving average model with exogenous variables (ARIMAX) revealed that both ARIMA and ARIMAX models fitted and estimated syphilis incidence well though ARIMAX model showed superior performance than the ARIMA (Zhang 2016). Li showed that the ARIMA-GRNN hybrid model was superior to the single ARIMA model in predicting the short-term TB incidence, especially in fitting and forecasting the peak and trough incidence (Li et al. 2019). Liu reported that both the ARIMA and BPNN models could be used to predict the seasonality and trend of pulmonary tuberculosis, but the BPNN model showed better performance (Liu et al. 2019).

This study will help in creating public awareness of future hypertension incidence and community health programs like routine door-to-door blood pressure checks, and sports to reduce inactivity could be incorporated. The study also provides a model to predict and undertake appropriate action to prevent, treat, and control hypertension. ARIMA has excellent predictive power as a univariate model. The ARIMA model used in this study can also be applied to other cases like the current pandemic, COVID-19, to sensitize policymakers to take pragmatic decisions and measures to improve health.

This study has several limitations. Firstly, the prevalence of hypertension may be underestimated due to the unawareness of blood pressure. Secondly, this is a univariate time-series study that is dependent on time only, but there could be other factors that might be attributed to the increase in hypertension cases. Thirdly, the short length from 2015-2019 of the time series may affect the modeling. Fourthly, the model should be used for short term predictions because long term predictions will just be the same as the mean of the series.

CONCLUSION

The model estimated an increase in hypertension cases for the next period, which was a critical input in managerial and administrative decision making. The forecast was accurate enough to allow for better planning and control than could be accomplished without the forecast.

Authors' contributions

All authors contributed to data analysis, drafting or revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Data availability

All data generated or analyzed during this study are included in this published article.

Funding

This study was funded by the National Key R&D Program of China (2017YFC0907000), Qing Lan Project of Jiangsu Province (2019), and Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). The funding agencies had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Ethical approval

This study was approved by the ethics committee of Nanjing Medical University. The study was conducted in accordance with the Declaration of Helsinki.

REFERENCES

- Anker D, Santos-Eggimann B, Zwahlen M, Santschi V, Rodondi N, Wolfson C, Chiolerio A (2020) Blood pressure control and complex health conditions in older adults: impact of recent hypertension management guidelines *J Hum Hypertens* doi:10.1038/s41371-020-0334-4
- Basu S, Millett C (2013) Social Epidemiology of Hypertension in Middle-Income Countries: Determinants of Prevalence, Diagnosis, Treatment, and Control in the WHO SAGE Study *Hypertension* 62:18-26 doi:10.1161/hypertensionaha.113.01374
- Box GEPAJ, G.Montrol, (1976) *Time Series Analysis: Forecasting and Control*, Revised Edition.
- GHS (2019) *The Ghana Health Sector (2019) report on Hypertension*
- Gu DF et al. (2005) Prevalence of cardiovascular disease risk factor clustering among the adult population of china - Results from the International Collaborative Study of Cardiovascular Disease in Asia (InterAsia) *Circulation* 112:658-665 doi:10.1161/circulationaha.104.515072
- He J et al. (2009) Premature deaths attributable to blood pressure in China: a prospective cohort study *Lancet* 374:1765-1772 doi:10.1016/s0140-6736(09)61199-5
- Le C, Zhankun S, Jun D, Keying Z (2012) The economic burden of hypertension in rural south-west China *Trop Med Int Health* 17:1544-1551 doi:10.1111/j.1365-3156.2012.03087.x
- Li YC, Wang LM (2012) Awareness, Treatment, and Control of Hypertension in China: Findings from 2010 China Chronic Disease and Risk Factor Surveillance *Heart* 98:E134-E134 doi:10.1136/heartjnl-2012-302920d.2
- Li Z et al. (2019) Application of a hybrid model in predicting the incidence of tuberculosis in a Chinese population *Infect Drug Resist* 12:1011-1020 doi:10.2147/IDR.S190418

- Liu Q et al. (2019) Forecasting the seasonality and trend of pulmonary tuberculosis in Jiangsu Province of China using advanced statistical time-series analyses *Infect Drug Resist* 12:2311-2322 doi:10.2147/IDR.S207809
- Mendis S, Johnston SC, Fan W, Oladapo O, Cameron A, Faramawi MF (2010) Cardiovascular risk management and its impact on hypertension control in primary care in low-resource settings: a cluster-randomized trial *Bulletin of the World Health Organization* 88:412-419 doi:10.2471/blt.08.062364
- Mendis S, Lindholm LH, Mancia G, Whitworth J, Alderman M, Lim S, Heagerty T (2007) World Health Organization (WHO) and International Society of Hypertension (ISH) risk prediction charts: assessment of cardiovascular risk for prevention and control of cardiovascular disease in low and middle-income countries *J Hypertens* 25:1578-1582 doi:10.1097/HJH.0b013e3282861fd3
- Mohsen Ibrahim M (2018) Hypertension in Developing Countries: A Major Challenge for the Future *Curr Hypertens Rep* 20:38 doi:10.1007/s11906-018-0839-1
- Stanley Jere and Edwin Moyo (2016) Modelling epidemiological data using Box-Jenkins procedure
- Wang Z et al. (2019) A national study of the prevalence and risk factors associated with peripheral arterial disease from China: The China Hypertension Survey, 2012-2015 *Int J Cardiol* 275:165-170 doi:10.1016/j.ijcard.2018.10.047
- Whelton PK et al. (2018) 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines *Circulation* 138:E484-E594 doi:10.1161/cir.0000000000000596
- WHO (2017) Non-communicable Diseases Progress Monitor 2017.
- WHO (2019) Hypertension
- Wilkins K et al. (2010) Blood pressure in Canadian adults *Health Rep* 21:37-46
- Zhang X, Zhang, T., Pei, J., Liu, Y., Li, X., & Medrano-Gracia, P., (2016) Time Series Modelling of Syphilis Incidence in China from 2005 to 2012 *PLoS one*, 11(2), e0149401. doi:doi:10.1371/journal.pone0149401

Figure and Tablelegends

Figure 1 Raw plots showing Monthly cases of hypertension over time

Figure 2 Graph showing decomposition of additive time series

Figure 3 ACF and PACF of hypertension cases

Figure 4 Multipleplots showing residuals of an ARIMA (1, 1, 0)

Figure 5 Forecasts of ARIMA (1, 1, 0) model

Table 1. Modeled data on hypertension cases