Box-Jenkins’ Methodology in Predicting Maternal Mortality Records from a Public Health Facility in Ghana

David Adedia¹, Salifu Nanga¹, Simon Kojo Appiah², Anani Lotsi³, Daniel A. Abaye¹*

¹School of Basic and Biomedical Sciences, University of Health and Allied Sciences, Ho, Volta Region, Ghana
²Department of Mathematics, College of Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
³Department of Statistics, University of Ghana, Legon, Ghana

Email: dadedia@uhas.edu.gh, snanga@uhas.edu.gh, skappiah@knust.edu.gh, alotsi@ug.edu.gh, *dabay@uhas.edu.gh

Abstract

The Millennium Development Goal (MDG) 5 advocated the reduction of maternal mortality rates significantly by 2015, however, maternal mortality rates continue to rise. Here, we modelled maternal mortality data for the years 2000 to 2013 obtained from a public hospital in Kumasi, Ghana. We applied the Box-Jenkins approach of univariate form of time series autoregressive integrated moving average (ARIMA). The output revealed that the ARIMA (1, 1, 1) model was most appropriate to model and predict monthly maternal cases with Akaike information criterion (AIC) value of 117.02 and Bayesian information criterion (BIC) value of 125.91. The Shapiro-Wilk normality test confirmed normality of the residuals. The Ljung-Box test on the residuals showed no serial correlation. The model was then validated based on the measures of accuracy. The results showed that the maternal mortality cases for the years 2000 to 2011 are high: minimum 3, median 11, mean 12 and maximum cases of 26 per month. The predicted mortality cases were 10 to 11 monthly for years 2012 to 2013, indicating that the target of MDG 5 could not be achieved by 2015. Fresh and perceptive strategies are urgently needed to arrest the unacceptably high death rates.

Keywords

Maternal Mortality, MDG 5, Box-Jenkins Methodology, ARIMA, Model Validation, Ghana

1. Introduction

Issues of maternal health have continuously received attention globally and na-
tionally since the 1980s. The UN Millennium Summit (2000) involving the UN member states, including Ghana, adopted Millennium Development Goals (MDGs) 8 which were geared towards the improvement of life of all people across the globe [1]. The MDGs have since been subsumed by the Sustainable Development Goals (SDGs) [2]. Maternal health improvement became the MDG 5 to be achieved by 2015. However, maternal mortality cases are still on the rise in Ghana [3]. Maternal health was brought into the international limelight by the thought provoking publication of Rosenfield [4]. The report showed that in many developing countries, maternal deaths were not considered as an important public health problem. The report indicated that mothers giving birth were one of the neglected health problems and this had resulted in the deaths of many women. And that, mortality rates for developing countries were 100 times more than in developed countries [4]. The report further stated that the programs that existed may not reduce the high maternal mortality rates recorded in these developing countries. Earlier, Harrison [5] also conducted a research and his analysis of 22,774 consecutive hospital births in Zaria, Northern Nigeria, showed the appalling mortality rates associated with child birth.

The deaths that occur in women during pregnancy or within 42 days after pregnancy termination are referred to as maternal mortality [5]. Obstructed labour, maternal hemorrhage, postpartum sepsis, eclampsia, unsafe abortion and anemia are among the listed causes of maternal mortality [3] [6]. Muchemi and Gichogo [7] estimated the maternal mortality ratio (MMR) for the world to be 210 per 100,000 live births and 480 per 100,000 live births for Africa in the year 2010. In contrast, the Centre in Charge of Maternal and Child Enquiries (CMACE, [8]) reported that maternal mortality in United Kingdom has dropped significantly ($P = 0.02$) from 13.95 per 100,000 maternities to 11.39 per 100,000 during the years 2003 to 2008.

Maternal mortality reduction is one of the MDG that Ghana seeks to achieve since it affects the development of the nation [9]. The objective of the MDG 5 is to improve maternal health, and to minimize maternal mortality (Target 6) ratio by 75% by 2015. Subjecting women to poor maternal health situation is also considered as a violation of their rights [10]. Globally, an estimated number of 289,000 women die annually and, 800 of these vulnerable women lose their lives daily from complications which are pregnancy-related [10]. According to WHO and UNICEF [11], the probability that a woman aged 15 years will die from maternal causes, is 1 in 3800 for developed countries compared with 1 in 150 for developing countries.

Despite interventions and several efforts by governments and other development partners towards achieving the goals of MDG 5, the MMR for developing countries still remains high [12]. In Ghana, maternal mortality is very prevalent, as women have a 1 in 68 lifetime risk of dying due to maternal causes [6]. Due to this high mortality rate, the government, policy makers and stakeholders are making efforts to introduce a number of policy interventions to help radically
reduce maternal mortality. In 2003, the government of Ghana passed the National Health Insurance (NHI) Act and implemented it in 2004 (The Act 650 of 2003 has since been amended by Act 852 of 2012 [13] [14]). Nationwide implementation commenced in 2004 via decentralized district-wide mutual health insurance schemes (DMHISs) [15] [16]. The NHI Act was introduced with the core aim of providing quality as well as affordable health services to all residents in Ghana [16] [17]. Women in Ghana, especially those in the rural areas, have limited access to health facilities and health personnel [18] and therefore, many nursing mothers and pregnant women receive health services from Traditional Birth Attendants (TBAs) [19]. Typically, TBAs are informally trained birth attendants whose skills are learnt and passed on from mother to daughter or niece [18].

Using the Box-Jenkins methodology [20], maternal deaths in the three northern regions of Ghana were assessed and found to be high [21]. Their findings also showed the seasonal nature of the maternal mortality in these regions of Ghana. Further, the authors found that except for the month of August, cases of maternal mortality increased from May to December, bringing to light the seasonality of the maternal mortalities. Maternal mortality trend analysis with Box Jenkins methodology has been carried out by many researchers (e.g. [21] [22] [23]). Box Jenkins methodology was also used to model malaria cases in Sudan [24], mortality due to malaria in Zambia [25], and cancer cases in Kenya [26].

In this report, we applied the principles of Box-Jenkins methodology to maternal mortality cases recorded at the Konfo Anokye Teaching Hospital (KATH), Kumasi, Ashanti Region, Ghana. The study seeks to model, validate and forecast the monthly maternal mortality at the hospital and highlight the trend of maternal mortality in the presence of programmes implemented to achieve the MDG 5. The findings of the study could serve as a guide for a review of MDG 5 with the passage of time and help assess the current interventions to curb the high numbers of maternal mortality.

2. Materials and Methods

2.1. Box-Jenkins Methodology

Box-Jenkins methodology is a statistical procedure that is used to model time series data by using autoregressive moving average (ARMA) or autoregressive integrated moving average (ARIMA) models [20]. The data is secondary monthly data, collected from KATH, Kumasi, Ghana which covered the period of 14 years, from 2000 to 2013 (See Supplementary Material). The data were modelled using ARIMA models. The ARIMA models help to fit datasets that have time series structure to describe the trend of maternal mortality and forecast points ahead within a given population. It also provides a forecast interval and it is based on a proven model. Forecasting methods can be divided into time series and explanatory types of analysis [27]. Time series models generally predict the continuation of historical patterns, such as employment rates, educational at-

DOI: 10.4236/ojapps.2018.86016
tainment, incidents of crime or disease burden within a community, if three basic conditions exist: 1) The availability of historical records or data, 2) The historical data are numerical and therefore, calculable, and, 3) The assumption that some portion of the past pattern will continue into the future.

2.2. The ARIMA Model

A dataset \( Y_t \) follows ARIMA model if the \( d \)th differences \( \nabla^d Y_t \) follows a stationary ARMA model. The parameters that help build the ARIMA model are three; \( p \), which determines the AR order; \( d \), denotes the number of differencing required before stationarity, and MA order is given by \( q \) \([28] [29]\).

Hence, ARIMA \( (p, d, q) \) is represented in a general form according to Tebbs \([28] [29]\) as:

\[
\nabla(B)(1-B)^d Y_t = \theta(B) e_t
\]

(1)

where, the AR and MA characteristic operators are

\[
\nabla(B) = 1 - \nabla_1 B - \nabla_2 B^2 - \cdots - \nabla_p B^d
\]

(2)

\[
\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q
\]

(3)

And,

\[
(1-B)^d Y_t = \nabla^d Y_t
\]

(4)

where,

\( \phi \) is the autoregressive parameter to be estimated; \( \theta \) is the moving average parameter to be estimated; \( \nabla \), the difference operator; \( B \), the backward shift operator; \( e_t \), random process having zero mean and variance not depending on time \( \sigma^2 \). Box and Jenkins \([20]\) proposed the estimation of parameters of ARIMA model, and their approach involves the steps: identification of ARIMA model, model parameter estimation, and model diagnostics \([28] [29]\).

2.3. Unit Root Tests

In order to make inferences in time series analysis, it is necessary to determine whether the time series is stationary or not. Prior studies have relied on Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test \([30]\) and the Augmented Dickey Fuller (ADF) test \([31]\) for assessing stationarity of the dataset \([32]\). The test assumes that \( y_t \) follows a randomness in the time series data:

\[
y_t = \rho y_{t-1} + e_t
\]

(5)

where, \( \rho \), the characteristic root of an AR polynomial and \( e_t \), white noise with mean zero and variance \( \sigma^2 \) \([28]\). The ADF test helps to test the null hypothesis of non-stationarity in the data. This results in the following ADF unit root test:

H\(_0\): \( \rho = 1 \) (non-stationary) versus H\(_1\): \( \rho < 1 \) (stationary).

Phillips and Perron \([33]\) introduced a robust PP test as an improved ADF test, and it counteracts the effects of general forms of heteroscedasticity and addresses issues of serial correlation. The KPSS test undertakes the hypothesis testing dif-
D. Adedia et al.

ferently and tests whether or not to reject the null hypothesis of stationarity [32].

2.4. Identification of ARIMA Model

There are techniques under ARIMA model identification which estimate the \( p \), \( q \) and \( d \) values. The autocorrelation function (ACF) and partial autocorrelation function (PACF) help to determine the \( p \), \( q \) and \( d \) values. The theoretical PACF of ARIMA \((p, q, d)\) process usually show non-zero PACF at first \( p \) lags, with remaining lags having zero PACF. The first \( q \) lags also report non-zero ACF and the remaining lags having zero ACF for the theoretical ACF. We determine \( q \) and \( p \) by the total frequency of the significant lags which are not zero for ACF and PACF respectively. If the values of \( p \), \( d \), and \( q \) are inaccurately selected, models derived can be inadequate, hence cannot be used for predictions [28][29].

2.5. Estimation of Model Parameters and Model Selection

If the ARIMA model is identified, then the maximum likelihood approach to estimating the parameters is used. In estimating the parameters, the log likelihood of a given \( p \), \( d \), \( q \) is maximized so that the probability of obtaining the observed data is maximized. Model estimation is followed by model selection, and it is done by considering minimum values of Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) [28] [29].

\[
\text{AIC} = -2 \ln(L) + 2h
\]

and

\[
\text{BIC} = -2 \ln(L) + \ln(n)h,
\]

where, \( L \) is the likelihood value of the likelihood function, \( h \) and \( n \) are number of parameters to be estimated and number of residuals respectively. For any two competing models, the model with the minimum AIC or BIC will be selected as a better one.

2.6. Model Diagnostic and Model Validation

Ljung-Box Q test is used to assess serial correlation of the residuals and it helps to determine the randomness of the residuals and model adequacy. Therefore,

\[
Q_m = n(n + 2) \sum_{k=1}^{m} (n-k)^{-1} r_k^2 \approx \chi^2_{m-1}.
\]

where, \( r_k^2 \) is the residuals autocorrelation at lag \( k \), \( n \) is the number of residual, and \( m \) is the number of time lags included in the test.

In this study, the level of significance is set to 5% and a model attains adequacy when the \( Q \) test statistics report \( p-value > 0.05 \). Absence of this renders the model inadequate and a better model should be identified and assessed. In addition, in order to achieve homoscedasticity, ACF and PACF will be plotted of the squared residuals. The Shapiro-Wilk test of normality and histogram will be used to assess the normality of the residuals. To validate the model selected, the dataset was modelled using a training set which comprised of data from 2000 to
2011 and validated using a testing sample from 2012 to 2013. The validation measures included root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} r_i^2}
\]

(9)

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |r_i|
\]

(10)

\[
\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{r_i}{o_i} \right|
\]

(11)

where \( r_i, n \) and \( o_i \) are the residuals, number of observations and the observed values respectively. The closer the validation measures for the errors from both models the better the training model.

Log transformation of data \((y_t)\) is among the family of methods called Box-Cox approach which is usually applied when there is high volatility in the data in order to stabilize the variance over time [34] [35]. For this study, a test statistic with \( p\)-value < 5% for any hypothesis testing was considered significant, which implied \( H_0 \) was rejected. The \( R \) statistical software was used for the analysis.

3. Results

3.1. Data Handling and Transformation

The data for this study is a time series (2000-2013) data from KATH, Kumasi on maternal mortality. The time plot in Figure 1 shows the maternal mortality at KATH, Kumasi from January 2000 to December 2011. The time plots coupled with computed mean and the variance showed that the data were volatile (large standard deviations) and therefore, log-transformed (Figure 2). Thus, the modeling was done using the log-transformed maternal mortality incidence data.

![Figure 1. Time series plot of maternal mortality cases at KATH, Kumasi, Ghana.](image)
3.2. Stationarity

The trend of the data over the years was assessed using the time plot. Examination of the dataset, revealed an existence of unstable trend. This was confirmed by computing the mean and variance of the dataset which revealed that the value for the variance was greater than that of the mean hence the instability of the data. This, consequently, led to the natural log-transformation of the data to stability.

The ADF, PP and KPSS tests were used to test for further stationarity. The results of the no differenced log data revealed that the ADF and PP test confirmed stationarity (Table 1). However, the KPSS test revealed otherwise. It should be noted that KPSS is the reverse of ADF and PP. The three tests confirmed stationarity when the series was differenced of the first order as can be seen in Table 1.

3.3. Model Identification

The output in Table 1 shows that after the first difference the dataset became stationary. By using the spikes in the ACF and the PACF plot of the log-differenced data of the first order, we suggest both the $q$ and $p$ values.

Figure 3 shows the ACF and PACF plots. The ACF plot has spikes at lags 0 and lag 1, which is the moving average part to the model and the PACF plot has spikes for lags 1, 2 and 3 which shows the autoregressive part. Therefore, models were tentatively suggested based on the combination of the significant spikes in both the ACF and PACF plots (Figure 3), and through Box-Jenkins approach the best model was selected as the best.

ARIMA (1, 1, 1* (Row 2)) in Table 2 is the best model because it is the model with the least AIC and BIC values.

3.4. Model Estimation

Table 3 contains the estimates of the ARIMA (1, 1, 1) model which shows a strong significance for the moving average component. This model will be used
Table 1. The results of the Unit Root tests.

<table>
<thead>
<tr>
<th>Different stationarity tests</th>
<th>The order of differencing</th>
<th>The test statistic values</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADF</td>
<td>0</td>
<td>−4.2998</td>
<td>0.01</td>
</tr>
<tr>
<td>PP</td>
<td>0</td>
<td>−11.270</td>
<td>0.01</td>
</tr>
<tr>
<td>KPSS</td>
<td>0</td>
<td>0.64742</td>
<td>0.02</td>
</tr>
<tr>
<td>ADF</td>
<td>1</td>
<td>−7.305</td>
<td>0.01</td>
</tr>
<tr>
<td>PP</td>
<td>1</td>
<td>−28.292</td>
<td>0.01</td>
</tr>
<tr>
<td>KPSS</td>
<td>1</td>
<td>0.012779</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 2. The fitted ARIMA (p, 1, q) models.

<table>
<thead>
<tr>
<th>Model</th>
<th>BIC</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ARIMA (1, 1, 0)</td>
<td>166.13</td>
<td>160.2</td>
</tr>
<tr>
<td>2* ARIMA (1, 1, 1)</td>
<td>125.91*</td>
<td>117.02*</td>
</tr>
<tr>
<td>3 ARIMA (2, 1, 1)</td>
<td>130.79</td>
<td>118.94</td>
</tr>
<tr>
<td>4 ARIMA (3, 1, 1)</td>
<td>133.51</td>
<td>118.69</td>
</tr>
<tr>
<td>5 ARIMA (2, 1, 0)</td>
<td>147.63</td>
<td>138.75</td>
</tr>
<tr>
<td>6 ARIMA (3, 1, 0)</td>
<td>143.74</td>
<td>131.89</td>
</tr>
</tbody>
</table>

Table 3. Model parameters estimates of ARIMA (1, 1, 1).

<table>
<thead>
<tr>
<th>Component</th>
<th>Coefficient</th>
<th>S.E</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.0496</td>
<td>0.0894</td>
<td>0.5791</td>
</tr>
<tr>
<td>MA(1)</td>
<td>−0.9504</td>
<td>0.0346</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Figure 3. ACF and PACF of first differenced log maternal mortality.

3.5. Model Diagnostic

From Figure 4, the residuals are the white noise with ACF having a spike at lag

\[ Y_t = (1 - 0.9504B)e_t = -0.9504e_{t-1} + e_t \] (12)
Figure 4. Diagnostic plot of residuals of ARIMA (1, 1, 1).

0, which may occur as a result of randomness. The Ljung-Box test also showed that the ARIMA (1, 1, 1) was adequate with p-value > 5% and could be used to forecast maternal mortality cases at the hospital.

The histogram (Figure 5) indicates a normality of the residuals and Shapiro-Wilk normality test reported a p-value of 0.158 > α = 0.05 which confirmed the normality of the residuals.

3.6. Model Validation

The dataset was partitioned as training and testing sample. The training sample contains about 85.7% (2000 to 2011) portion of the dataset for modeling the data. The sample for testing the validity of the model (test sample) contains the remaining portion, 14.3% (2012 to 2013) of the dataset. Based on the estimates of RMSE, MAE and MAPE for both training model and testing model in Table 4, it is demonstrated that the training model has a good predictive ability since estimates from both models are close. The closeness in errors from both models confirms the fact that the training model can be used for prediction. Figure 6 shows the predicted maternal mortality cases which lie within the 95% confidence intervals. The lower confidence limit (LCL) and upper confidence limits (UCL) are also indicated. Mortality cases are expected to be constant over the time period from 2012 into 2014 for the hospital.

4. Discussion

Our study confirms that, instead of maternal mortality cases declining as being sought by Target 6 of the MDG 5, there were increases in death rates at the KATH, Kumasi, Ghana over the years to 2013. This study has revealed that maternal mortality rates are expected to be on a constant trajectory over the years 2012 to 2013 even into 2014 if the prevailing conditions remain from the previous years. These observations have also been made in other studies [3] [12].
Figure 5. Histogram of the residuals.

Figure 6. 24 months forecasts of maternal mortality. LCL is the lower confidence limit and UCL is upper confidence limits for the forecast values.

Table 4. Model validation.

<table>
<thead>
<tr>
<th>Model Fit Indexes</th>
<th>Training Model</th>
<th>Testing Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.3527</td>
<td>0.2583</td>
</tr>
<tr>
<td>MAE</td>
<td>0.2706</td>
<td>0.2016</td>
</tr>
<tr>
<td>MAPE</td>
<td>13.5079</td>
<td>9.5522</td>
</tr>
</tbody>
</table>

Our study supports the report by Commonwealth Health Online [36], that the rate of maternal mortality in Ghana continues to be high, and that, the MDG 5 would not be achieved by 2015. This has negative implications on the Ghanaian society.

According to the UN Agencies report [37], championed by the Maternal Mortality Estimation Inter-Agency Group (MMEIG), from 1990 to 2013, there were reductions in maternal mortality cases, that is, the maternal mortality rate decreased from 760 to 380 cases for every 100,000 live births in Ghana. However, nationally, Ghana recorded 3100 maternal deaths in 2013. A lot more effort and new strategies should be implemented before Ghana could achieve the MDG 5. Based on our findings, maternal mortality of 10 to 11 cases per month are unacceptably high at KATH, Kumasi, compared with the Northern Region of Ghana.
which recorded 66 cases for the whole year of 2014 [37]. That is, there are five to six deaths per month in the Northern Region, Ghana. Secondly, the sum of mortality cases of 10 to 11 every month is almost the same as the number of mortality cases that the whole country is expected to have (185 deaths per every 100,000 live births), and this is not in favor of the target of MDG 5 which seeks to achieve a mortality of less than 185 cases for every 100,000 live births by 2015. Unfortunately, we were unable to access the mortality rates at the hospital from 2012 to 2015 in order to make final comparisons.

5. Conclusion

This study applied the Box-Jenkins methodology to model the maternal mortality cases recorded at KATH, Kumasi, Ghana, using data from 2000 to 2013. The time series modeling was employed by first assessing the time plot, ACF and the PACF of the series. The time plot showed fluctuations in mortality from 2000 to 2011, with 2011 recording the highest mean maternal mortality of 12 cases. The dataset was natural log transformed because it was volatile. Finally, the appropriate model ARIMA (1, 1, 1) was used to forecast two years (24 months) for the maternal mortality cases at KATH, Kumasi. The model adequacy and validation have also shown to be appropriate in predicting the maternal mortality cases, and was used to forecast data from 2012 to 2013. The forecast values fell within the required 95% confidence interval highlighting the adequacy of the fitted model. The results of the forecasting showed that from 2012 to 2013, the maternal mortality rates were stable, and were estimated to be 10 to 11 cases monthly. These predicted monthly maternal mortality cases are unacceptably high and this is not in favour of the target of MDG 5. These findings could serve as a guide for a review of MDG 5 and help scrutinise the on-going interventions to curb maternal mortality. The MDG 5 was not achieved by the set time of 2015.

Acknowledgements

The authors would like thank the KATH, Kumasi, Ghana, for allowing access to the secondary data.

Conflicts of Interests

We declare no conflicts of interests.

References


https://jamanetwork.com/journals/jama/article-abstract/378024
https://doi.org/10.1001/jama.1989.03430030064035


https://doi.org/10.11604/pamj.2014.17.201.3694


[10] UN Human Rights, OHCHR. 


https://doi.org/10.1596/978-0-8213-6949-4

https://doi.org/10.1093/heapol/czn002

https://www.academia.edu/1873350/

https://doi.org/10.4236/ojs.2016.62025
http://people.stat.sc.edu/Tebbs/stat520/f13notes.pdf
http://www.bankofengland.co.uk/education/Pages/ccbs/handbooks/ccbshb22.aspx
[34] Bartczak, A., Glazik, R. and Tyszkowski, S. (2014) The Application of Box-Cox Transformation to Determine the Standardised Precipitation Index (SPI), the Standardised Discharge Index (SDI) and to Identify Drought Events: Case Study in

