International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Autoregressive Integrated Moving Average Model for Predicting Trend of Injury Mortality in India

M J Thirunavukkarasu[1], Bikash Kumar Das[2], H N Vrushabhendra[3]

[1] Assistant Professor cum Statistician, Department of Community Medicine, Sri Venkateshwaraa Medical College Hospital & Research Centre, Ariyur, Puducherry, India

[2] Associate Professor, Department of Community Medicine, Sri Venkateshwaraa Medical College Hospital & Research Centre, Ariyur, Puducherry, India

[3] Professor and Head , Department of Community Medicine, Sri Venkateshwaraa Medical College Hospital & Research Centre, Ariyur, Puducherry, India

Abstract: ***Objective: Injury mortality is presently an increasing public health problem in India. Reducing the loss due to injuries has become a most important priority of public health policies. Early caution of injury mortality based on surveillance information is essential for controlling the disease burden of injuries. We conducted this study to find the possibility of applying autoregressive integrated moving average (ARIMA) models to predict injuries mortality in India. Method: The yearly injuries mortality data in India (2000 to 2015) were used to fit the ARIMA model. The Ljung-Box test was used to measure the 'white noise' and residuals. The mean absolute percentage error (MAPE) between observed and fitted values was used to evaluate the predicted accuracy of the constructed models. Results: A total of 993912 injury-related deaths in India were identified during the study period; the average mortality rate was 62119.50, SD 11313.468, minimum 44909 and maximum 78600 persons. This ARIMA (1, 1, 2), passed the parameter (p<0.01) and residual (p>0.05) tests, with high*** R^2 ***and low RMSE, MAPE, MAE, and NBIC 0.99, 479.296, 0.551, 315.131, 13.067 respectively. Conclusions: The ARIMA (1, 1, 2) model could be applied to predict mortality from injuries in India.***

Keywords: Injuries, mortality, epidemiology, caution criteria, India

1. Introduction

Injuries that affect all ages of the population have become a serious worldwide public health threat. Deaths caused by injuries have a serious impact on communities and families.[1] The World Health Organization (WHO) and the Global Burden of Diseases Study (GBD) suggest that injuries account for 3.9 million deaths worldwide [2], of which about 90% occur in low- and middle-income countries. The majority of these deaths are attributable to road traffic injuries, falls, drowning, poisoning and burns [2]. In 2004, WHO estimated about 0.8 million deaths in India were due to injuries [2]. Direct Indian estimates of unintentional injury deaths relying on annual National Crime Records Bureau (NCRB) reports of injury deaths from police records showed only 0.3 million injury deaths in 2005 [3], but police record are subject to under-reporting and misclassification [4–6]. According to the latest report from the WHO, approximately 5.14 million people died from injuries in 2012, an incidence of 727 per million persons [7]. The autoregressive integrated moving average (ARIMA) model, one of the most classical methods of time series analysis, was first proposed by Box-Jenkins in 1976[8]. It is represented as a moving average (MA) model combined with an autoregression (AR) model to fit the temporal dependence structure of a time series using the shift and lag of historical information. In epidemiology, this model has been widely used to predict the incidence of infectious diseases such as dengue fever [9], avian influenza H5N1 [10] and hepatitis E in [11]. Predicting the number of deaths due to injuries in future years will generate useful information for designing the strategies of public health services. The objective of this study was to describe the temporal trends of injury mortality in India and to determine the possibility of applying ARIMA models to forecast injury mortality in the future years.

2. Materials And Methods

The present study was a time-series data on Injury of mortality for the year from 2000 to 2015 in India has been collected from the web site www.who.int/tb/data maintained by the Department of The World Health Organization (WHO) [12]. The data was analysed using Statistical Package for Social Sciences (SPSS version 23 as it is licensed with the SVMCH & RC) and fit the best suitable ARIMA model for the Injury mortality data. The performance criteria was used to determine if the model was correctly specified. Forecasting of the Injury of mortality was also done using the best fit.

3. Model Fitting

The Box–Jenkins methodology was adopted to fit the ARIMA (p, d, q) model. Before constructing the model, we have to identify the stationary state of observed data in in the series, of which the mean value remains constant. If non-stationary, the data would be transformed into a stationary time series by taking a suitable difference. The Ljung–Box test was used to measure the 'white noise' and residuals in the study. To determine the ARIMA model three steps were performed; model identification, parameter estimation and testing, and application. The orders of the model were identified initially by the cut-off figure of the autocorrelation function (ACF) and the decay figure of the partial ACF (PACF). Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Normalized Bayesian information criterion (NBIC) was used

to select an optimal model; the less error value is better fit of the data. The conditional least-squares method was used for parameter estimation, and the *t* test was used for parameter testing. A parameter without statistical significance had to be removed from the model. Following formula was computed for identify error value

$$RMSE = \left(\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 / n \right)^{1/2}$$

$$MAPE = (1/n) \sum_{t=1}^{n} \left| Y_t - \hat{Y}_t \right| / \hat{Y}_t \times 100, \text{and}$$

$$MAE = \sum_{i=1}^{n} \left| (Y_i - \hat{Y}_i) \right| / n, \text{ where } \hat{Y}_t \text{ the actual Injury}$$

mortality value, and Y_t the predicted Injury mortality value and n is the number of years used as a prediction period. NBIC(p,q) = In v^* (p,q) + (p+q) [In(n) /n] ; where p and q are the order of AR and MA processes respectively and n is the number of observations in the time series and v^* is the estimate of white noise variance σ^2. Finally, the fitted model was applied to forecast injury mortality in future years.

4. Statistical Analysis

The result of trend is revealed that the series of yearly injury mortality data in India from 2000 to 2015 was decreasing trend with a non-stationary sequence. Figure 1 and 2 shows that the auto-correlation function and partial auto-correlation function of the non-stationary mortality rate of Injury.

From these figures two facts have been emerged out, first the ACF declines very slowly then ACF up to 14 lags are individually statistically different from zero or, they all were found to lie outside the 95% confidence bound and secondly, after the first lag the PACF dropped dramatically and all the PACF after lag 1 were found to be non-significant. The correlogram represents that ACF remain close to 1 throughout, declining to zero gradually. So it is expected that high negative auto-correlations exists indicating decreasing in mean. When the variance of the realization appears fairly stable, the means are definitely decreasing and this series has an downward trend. It is regarded as non-stationary in mean.

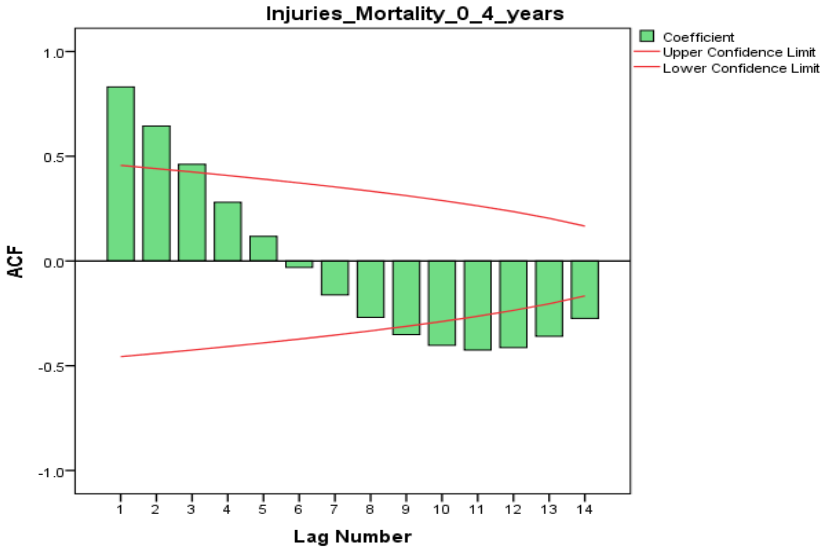

Figure 1: Plot of ACF for Injury mortality rate in India

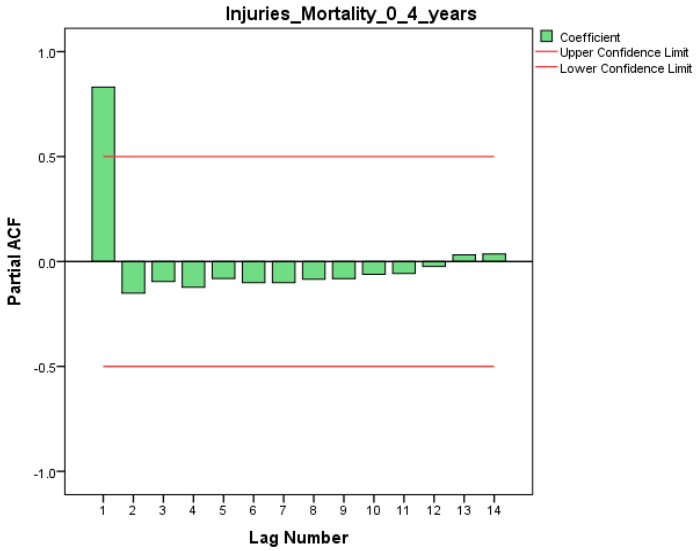

Figure 2: Plot of PACF for Injury mortality rate in India

The first-order differentiation to stabiles the variances. After first-order differentiation (d = 1), the data that were not 'white noise' ($p<0.01$) were dispersed horizontally around zero, suggesting that they were stationary. The ACF and PACF for first-order differentiated data are shown in figure 3 and 4. From these figures it was clear that at the first difference of injury mortality series was found to be more stable variance than the original injuries series. Tentatively selected fifteen ARIMA models at different values of p , d, and q were estimated and comparison among family of different parametric combinations of ARIMA (p,d,q) was done according to the minimum values of Normalize BIC,RMSE, MAPE, MAE and maximum value of R^2 which are given in Table1.

From the table 1 it is clear that among the different ARIMA model the ARIMA (1,1,2) had the higher value of R^2 with minimum values of Normalize BIC, RMSE, MAPE, MAE in comparison to that of the other model and the tentative model for the Injury mortality data set was ARIMA(1,1,2) in table 2.

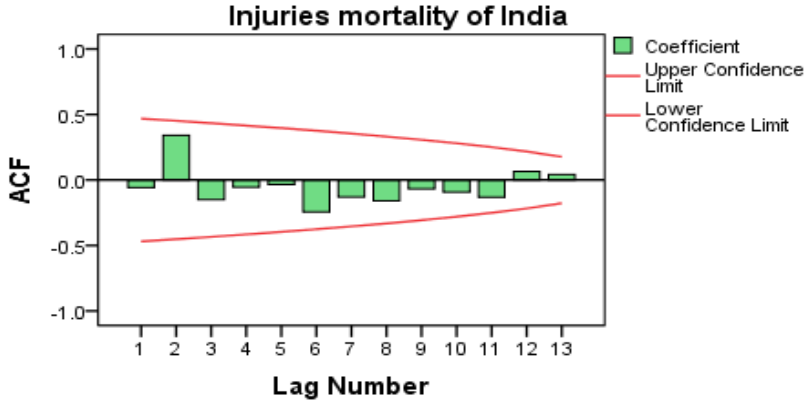

Figure 3: ACF plot of the first differenced Injury mortality

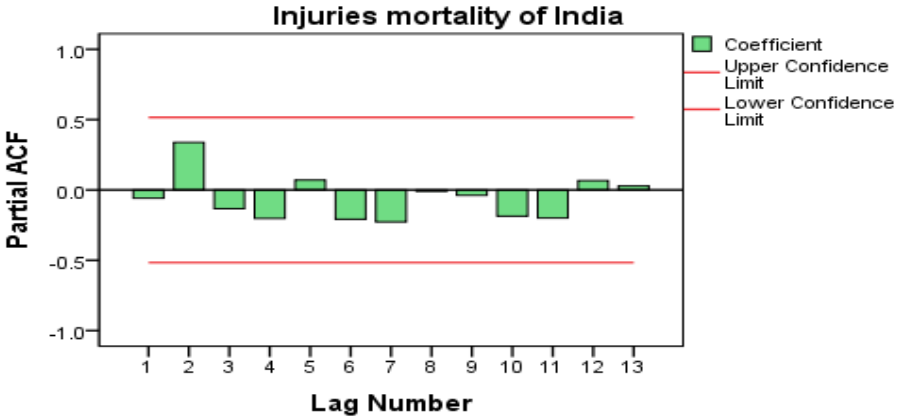

Figure 4: PACF plot of the first differenced Injury mortality

Table 1: Summary statistics of the different smoothing parameters and model performance measures

Model ARIMA(p,d,q)	Values of model selection criteria				
	R^2	RMSE	MAPE	MAE	NBIC
ARIMA(0,1,0)	0.99	590.886	0.788	422.996	12.944
ARIMA(0,1,1)	0.99	612.267	0.788	425.303	13.195
ARIMA(0,1,2)	0.99	485.348	0.588	318.605	12.911
ARIMA(1,1,0)	0.99	610.938	0.789	429.019	13.191
ARIMA(1,1,1)	0.99	606.326	0.747	416.274	13.356
ARIMA(1,1,2)	**0.99**	**479.296**	**0.551**	**315.131**	**13.067**
ARIMA(2,1,0)	0.99	535.89	0.652	365.171	13.109
ARIMA(2,1,1)	0.99	552.727	0.657	365.867	13.352
ARIMA(2,1,2)	0.99	483.133	0.544	311.279	13.263
ARIMA(3,1,0)	0.99	538.883	0.648	358.169	13.301
ARIMA(3,1,1)	0.99	489.971	0.534	302.893	13.291
ARIMA(3,1,2)	0.99	496.929	0.552	315.053	13.5
ARIMA(4,1,0)	0.99	497.190	0.543	309.822	13.321
ARIMA(4,1,1)	0.99	517.898	0.541	308.628	13.583
ARIMA(4,1,2)	0.99	530.964	0.507	296.866	13.813

Table 2: Estimates of ARIMA (1,1,2) model for Injury mortality rate in India

Coefficients	Estimates	Std.Error	t-value	p-value
Constant	2237.083	165.849	-13.489	0
AR1	0.802	0.503	-1.503	0.03
MA1	0.754	0.097	7.772	0
MA2	0.997	40.756	-0.024	0.01

It is clear from the table 2, that all the parameter estimates were found to be highly significant. Since all the model selection criterion measures were found to be minimum and the coefficient determination was 0.99 which means that 99% of variation in the data series was explained by the ARIMA (1, 1, 2) model.To check the auto-correlation assumption, "Box-Ljung" test was used, it was found that the $\Pr(|\chi_1^2| \geq 0.0724)=0.7879$, which strongly suggested that the acceptance of no auto-correlation among the residuals of the fitted ARIMA (1,1,2) model at 5% level of significance. The ACF and PACF of the residuals of ARIMA (1, 1,2) model is depicted in the figure 5. As all the

ACF and PACF values were found to be within confidence bound the residuals due to ARIMA (1,1,2) model possesses white noise.

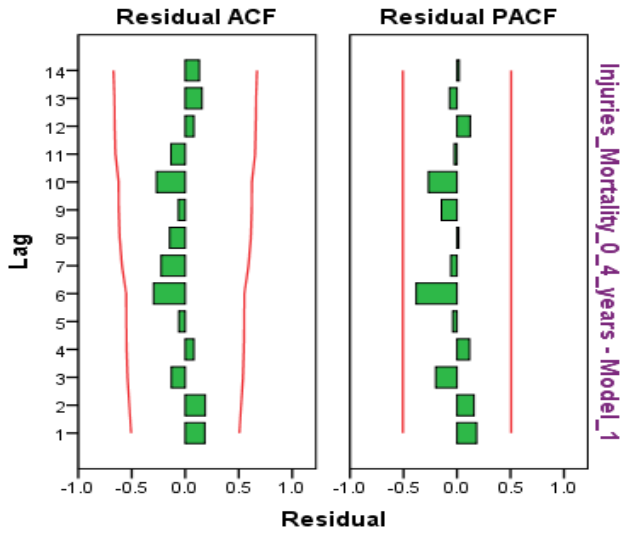

Figure 5: Residual plot of ACF and PACF of fitted ARIMA (1 ,1, 2)model

Again, to check the normality assumptions, "Jarque-Bera" test was used. From the test, it was found that the $Pr(|\chi_2^2| \geq 0.0097)=0.9951$, which strongly suggested the acceptance of the normality assumption that the residuals of the fitted ARIMA (1,1,2) model for Injury mortality series followed normal distributions. Finally, considering all graphical and formal test, it was clear that the fitted ARIMA (1,1,2) model was found to be most appropriate one among the ARIMA stochastic models employed to the Injury mortality rate in India and trend is predicted from 2000 to 2020 in the figure 6 and table 3.

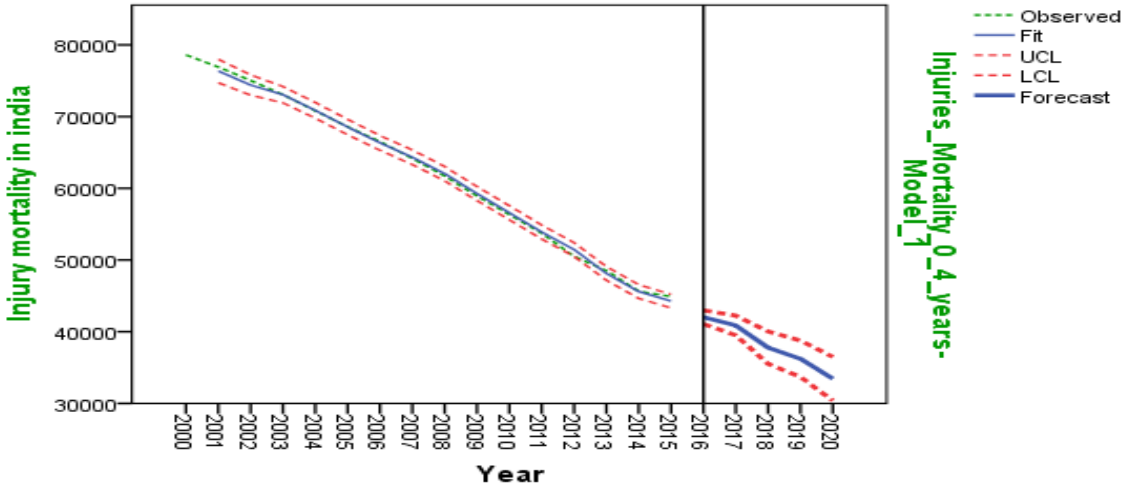

Figure 6: Trends in Injury mortality of India based on ARIMA (1, 1, 2) model

Table 3: Forecasted Injury mortality rate with 95% confidence interval

Year		2016	2017	2018	2019	2020
ARIMA (1,1,2)	LCL	41081	39500	35509	33685	30387
	Prediction	42029	40870	37768	36225	33431
	UCL	42978	42241	40028	38765	36476

UCL =Upper Confidence Limit, **LCL**= Lower Confidence Limit

5. Conclusions

The government urgently needs to evaluate the loss caused by injuries, a stochastic model accounts for patterns in the past movement of a variable and uses that information to predict its future injury mortality movements. To select the best model for a particular time series the latest available model selection criteria are used. The study revealed that ARIMA (1,1,2) models are appropriate for Injury mortality in India respectively and it is to be noted that the short-term forecast is better as the error of forecast increases with the increase of the period of forecast. Our modelling approach shows that applying the ARIMA time series models to forecast injury mortality in India is feasible and historical surveillance data are important tools for monitoring and forecasting injuries.

Ethical approval: Not required.

References

[1] Takala J, Hamalainen P, Saarela KL, et al. Global estimates of the burden of injury and illness at work in 2012. J Occup Environ Hyg, 2014; 326 – 37.
[2] World Health Organization: The Global Burden of Disease: 2004 update. Geneva: World Health Organisation; 2008.
[3] National Crimes Records Bureau: Accidental deaths and suicides in India, 2005. New Delhi: Government of India; 2006.
[4] Dandona R, Kumar GA, Ameer MA, Reddy GB, Dandona L: Under-reporting of road traffic injuries to

the police: results from two data sources inurban India. Inj Prev 2008, 14:360–365.
[5] Sanghavi P, Bhalla K, Das V: Fire-related deaths in India in 2001: a retrospective analysis of data. Lancet 2009, 373:1282–1288.
[6] Gururaj G, Sateesh V, Rayan A, Roy A, Amarnath, Ashok J: Bengaluru injury/road traffic injury surveillance programme: a feasibility study. Bengaluru: National Institute of Mental Health and Neuro Science; 2008.
[7] World Health Organization. Health statistics and information systems/Estimates for 2000-2012/causespecificmortality.2014. http://www.who.int/healthinfo/global_burden _disease /estimates/ en/index1.html (accessed 20 Sept 2014).
[8] Box, G.E.P., and G.M. Jenkins, 1976. Time Series Analysis and Control, Revised Ed, Holden Day.
[9] Earnest A, Tan SB, Wilder-Smith A, et al. Comparing statistical models to predict dengue fever notifications. Comput Math Methods Med 2012;2012:758674.
[10] Kane MJ, Price N, Scotch M, et al. Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. BMC Bioinformatics 2014;15:276.
[11] Ren H, Li J, Yuan ZA, et al. The development of a combined mathematical model to forecast the incidence of hepatitis E in Shanghai, China. BMC Infect Dis 2013;13:421.
[12] (2015) WHO TB Report 2014. http://www.who.int/tb/publications/global_report/en/