Int J Epidemiol Health Sci 2024;5: e64
doi: http//doi.org/10.51757/IJEHS.5.2024. 709969

Original article

Forecasting the trend of road traffic crashes and their accompanying fatalities in Nigeria (1960–2020)

Haruna Umar Yahaya, Mohammed Tanimu *, Olisaemeka Obi

Department of Statistic, University of Abuja, Abuja, Nigeria.

***. Corresponding author:** Mohammed Tanimu, Department of Statistic, University of Abuja, Abuja, Nigeria. Email: mohammed.tanimu@uniabuja.edu.ng.

Cite this article: Yahaya, H.U., Tanimu, M., Obi, O. Forecasting the trend of road traffic crashes and their accompanying fatalities in Nigeria (1960–2020). Int J Epidemiol Health Sci 2024;5:e64. Doi: 10.51757/IJEHS.5.2024.709969.

Abstract

Background: Despite efforts to improve road safety, Nigeria continues to have a high number of traffic crashes and fatalities. This has contributed to Nigeria's death rates. The study aimed to model and anticipate the trend of road traffic collisions and fatalities in Nigeria.

Methods: The study employed a descriptive retrospective approach to examine the trend of road traffic crashes and their associated fatalities in Nigeria. The study used secondary data from the Federal Road Safety Corps (FRSC) database spanning sixty years, from 1960 to 2020.

Results: Between 1960 and 2020, Nigeria had an average of 19014 road traffic collisions and 6104 fatalities. The number of road traffic crashes in Nigeria increased from 1961 (10963) to 1976 (40881), then began to fall to (9694) in 2020, although road traffic fatalities continue to rise somewhat. The ARIMA (1,1,0) and Random walk models were shown to be the best fitted time series models for predicting the number of crashes and associated fatalities.

Conclusion: Trend analysis in road traffic accidents remains an important component of ongoing efforts to minimize fatalities and injuries while promoting safer and more sustainable transportation systems. This study will investigate and synthesize current trends in road traffic accidents and fatalities, giving light on the factors that influence the road safety landscape.

Keywords: Random Work Model**,** ARIMA Model, Road Traffic Crashes, Road Traffic Fatalities, Nigeria

Introduction

Road traffic accidents and fatalities are a major public health concern in Nigeria, with severe social and economic consequences. According to the World Health Organization (1), Nigeria has one of the world's highest rates of traffic fatalities, with an estimated 33.7 deaths per 100,000 people in 2018. This is nearly triple the global average of 18.2 fatalities per 100,000 people. Despite efforts to improve road safety, Nigeria continues to have a high number of traffic crashes and fatalities. Nigeria's road safety condition is characterized by inadequate infrastructure, bad road design, a failure to enforce traffic laws, and poor driving behavior. Furthermore, the growing number of automobiles on Nigerian roads, as well as the rapid expansion of urban centers, have compounded the country's road safety situation.
The rising number of road traffic incidents and fatalities in the country is a major public health concern that necessitates statistical research to better understand the underlying causes and develop effective solutions. The lack of statistical analysis hampers our ability to reach correct conclusions and make evidence-based policy recommendations. Thus, the goal of this study was to conduct a rigorous

statistical analysis of the trajectory of road traffic crashes and associated fatalities in Nigeria, utilizing advanced statistical approaches to uncover the causes driving the increase and construct models that can forecast future trends. The statistical analysis will shed light on the causes of road traffic collisions and fatalities in Nigeria, as well as inform evidence-based policy measures that can reduce the number of accidents and fatalities in the country.

Road traffic accidents and fatalities are a major public health concern worldwide. According to the World Health Organization (1), road traffic accidents cause a significant fatality rate, notably in South-Western Asia (2). Road traffic injuries are a big issue in China, with a high number of fatalities reported by national data sources (3). Similarly, in Qatar, road traffic accidents and injuries are a major health concern and the top cause of mortality (4). The consequences of road traffic accidents are not limited to certain areas. In India, numerous fatalities and injuries were reported in a single year, emphasizing the gravity of the situation (5). Sri Lanka has similarly seen a long-term increase in road traffic collisions, injuries, and fatalities during the last few decades (6). These findings highlight the global character of the problem and the importance of effective responses. Traffic infraction behaviors are one factor that contributes to road traffic crashes. According to studies, traffic offenses are a major source of crashes and fatalities in a number of nations, including China, Europe, and the United States (7). Understanding and correcting these violations can play an important role in lowering the number of road traffic accidents. The consequences of road traffic accidents go beyond the acute health impact. Road traffic accidents, for example, have substantial economic effects in Ghana and Nigeria, where a large number of deaths and injuries have been reported over the last decade (8,9). Pedestrians and spectators are especially vulnerable to death in road traffic incidents (10, 11). This emphasizes the importance of comprehensive policies that address all of the elements that contribute to road traffic crashes, such as infrastructure, driver behavior, and public awareness. To summarize, road traffic accidents and fatalities are a global public health concern. The literature review focuses on the occurrence of road traffic injuries in various nations and areas, highlighting the need for effective treatments. Understanding the epidemiological trends, contributory factors, and consequences of traffic crashes is critical for creating evidence-based prevention and mitigation techniques.

Materials and methods

The study employed a descriptive retrospective approach to examine the trend of road traffic crashes and their associated fatalities in Nigeria. The study used secondary data from the Federal Road Safety Corps (FRSC) database spanning sixty years, from 1960 to 2020. The analysis was restricted to data from the FRSC database due to the agency's responsibility to collect and maintain data on road traffic crashes in Nigeria. The data contained information on the date, time, location, number of cars involved, and number of casualties caused by road traffic accidents. The FRSC personnel collected the data in a uniform format, which they then entered into the agency's database.

The study employed descriptive statistics to examine the trend of road traffic accidents and fatalities in Nigeria. To summarize the data, statistical methods such as frequency distributions, percentages, and measures of central tendency were used. The study also used time-series analysis to look at the trends in road traffic crashes and fatalities over a sixty-year period. Descriptive statistics were used to summarize the data on the number of road traffic crashes and fatalities in Nigeria from 1960 to 2020. The analysis was carried out using Econometric views and Statgraphics Software.

The ARIMA Model

ARIMA is an acronym for Auto-Regressive Integrated Moving Average. This is a recognized time series model, which could be defined algebraically as:

$$Y_t = \mu + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{p-1} + e_t - \delta_1 et_{-1} + \delta_q e_{t-1} \quad \textbf{\textit{Formula 1}}$$

at time t =1,…,n, where et-j (j=0,1,…,q) are the lag forecast errors. The p + q+1 unknown parameters μ, αj …,αp and αi …,αq are typically calculated by minimizing the squared residuals (12, 13).

Using the ARIMA technique, the dependent variable yt is predicted in the first half of the right-hand side of equation (1) above based on its values in previous time periods. This is the autoregressive (AR) component in equation (1) above. In the second component, the dependent variable yt is also affected by the residual values from previous time periods, which can be seen as preceding random alarms. This is the moving average (MA) part of equation (1).

In addition to the AR and MA parameters, ARIMA models may have a constant. The interpretation of a statistically significant constant is determined on the model used. Two example circumstances are:

i. There are no autoregressive parameters in the series. In this situation, the predicted value of the constant is the average of the series;

ii. ii. The scenario with autoregressive parameters in the series. In such cases, the constant denotes the intercept. If the series is different, the constant reflects the series' mean or intercept. In the non-seasonal situation, the simple ARIMA (p, d, q) model is employed, with p denoting the number of autoregressive terms, d the number of non-seasonal differences, and q the number of lagged forecast errors in the prediction equation. However, climate data frequently includes seasonal variations. Thus, it is more likely to incorporate the entire Seasonal Auto Regressive Integrated Moving Average (SARIMA) model.

SARIMA $(p\ d\ q)(P\ D\ Q)_S$ ***Formula 2***

with P denoting the seasonal AR-model, D the seasonal differencing, and Q the seasonal MA-model. The subscript s represents the number of periods in the season. Mathematically, the generic version of the model expressed in equation (2) above can be stated in the backshift notation (B):

$$\alpha_{AR}(B)\ \alpha_{SAR}(B^s)(1-B^s)^d(1-B^s)^D\ y_t = \delta_{MA}(B)\delta_{SAM}(B^s)e_t$$ ***Formula 3***

where αAR is the non-seasonal AR parameter, δMA the non-seasonal MA parameter, αSAR the seasonal AR value, and δSAM the seasonal MA parameter.

The Stationarity Condition

Stationarity is an essential requirement for ARIMA models. In practice, the mean and variance should be constant as a function of time prior to conducting the study. Otherwise, previous impacts would accrue and the values of successive yts would approach infinity, rendering the process non-stationary. To detect first-order non-stationarity in ARIMA models, differentiate the observations d times and use Δd yt instead of yt as the time series. This is normally done through the metamorphosis:

$$\Delta y_t = y_t - y_{t-1}$$ ***Formula 4***

The operations of equation (4) yield the values d = 0,1,2,... for the non-seasonal component and D=0,1,2,... for the seasonal part, which serve as a suggestive aid in eliminating the first order nonstationary in the model identification process.

It should be noted that in the case of second order non-stationarity, a simple transformation (such as the log transformation) may be a desirable operation to execute when found.

Applying the ARIMA technique

Previously, we focused on the Box and Jenkins methodologies and their benchmark model application approach, which consists of three steps:

(a) identification.

(b) Estimation and

(c) Forecasting/diagnostic checks. During the identification stage, the linear least squares approach is used to calculate preliminary values for the p, d, and q sets. During the identification stage, a stationary or weakly stationary condition is obtained by differencing and transforming the data, if appropriate. The ACF and PACF charts are then used to generate alternative models by finding the orders p and q in the Seasonal ARIMA (p, d, q,) (P, D, Q,) S model. The goodness of the best models can be assessed using the Mean Square Error (Residuals) MSE or the Akaike Information Criterion AIC (14,15).

Autocorrelation and partial autocorrelation

Autocorrelation and partial autocorrelation are time series techniques for determining the linear relationship between lagged values of time series. The greater the departure of these coefficients from zero, the more dependent the series is at a given point in time on its lag values. Correlograms (Plots of Autocorrelation Functions (ACF) and Partial Autocorrelation Functions (PACF)) graphically represent the magnitude of the series' dependency on past values (12).

Forecasting evaluation

Forecasting Evaluation Criteria Numerous error measures are available for forecast evaluation; thus, this study evaluates the forecasting ability of state space and Box-Jenkins type models by means of three different loss functions. These are root mean squared error (RMSE), mean absolute error (MAE) and Theil's U statistic which are defined as follows:

$$\boldsymbol{RMSE} = \sqrt{\boldsymbol{MSE}} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\mathrm{A}_t - \mathrm{F}_t)^2}$$ ***Formula 5***

$$\boldsymbol{MAE} = \frac{1}{n}\sum_{t=1}^{n}|(\mathrm{A}_t - \mathrm{F}_t)^2|$$ ***Formula 6***

$$Ut = \frac{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\mathrm{A}_t - \mathrm{F}_t)^2}}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\mathrm{A}_t - A_{t-1})^2}}$$ ***Formula 7***

Where At is the actual value in time t, and Ft is the anticipated value in time t. Theil's U statistic evaluates forecast accuracy across models. The overall performance of the estimating methods was accessed using the average of the three loss functions, that is $Average = (RMSE + \boldsymbol{MMQ} + Ut)/3$, the approach with the least Average is the best.

Table 1. Descriptive Statistics

	Number of crashes	*Number of killed*
Count	61	61
Average	19013.7	6103.61
Standard deviation	8344.39	2624.18
Coeff. of variation	43.8862%	42.9939%
Minimum	8477.0	1083.0
Maximum	40881.0	11382.0
Range	32404.0	10299.0
Stnd. skewness	2.59822	-0.599371
Stnd. Kurtosis	-0.414571	-1.14589

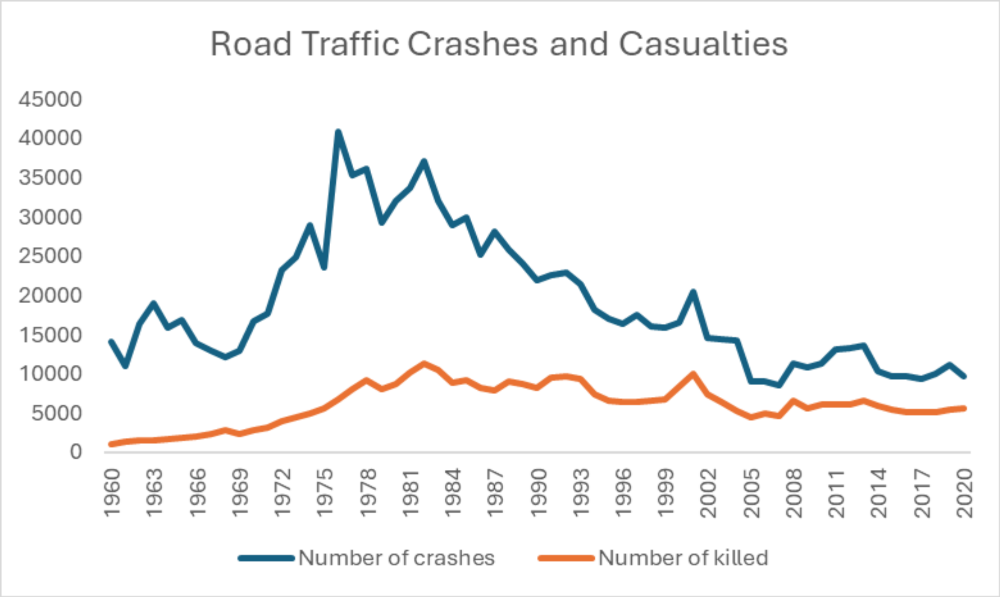

Figure 1. Trend of road traffic crashes and fatalities in Nigeria from 1960 to 2020

Table 2. Unit Root Test

		Number of crashes	**Number of killed**
		t-Statistic	**t-Statistic**
Augmented Dickey-Fuller test statistic		-10.68990	-6.606606
Test critical values:	**1% level**	-3.546099	-3.546099
	5% level	-2.911730	-2.911730
	10% level	-2.593551	-2.568766
	Prob.*	0.0000	0.0000
*MacKinnon (1996) one-sided p-values.			

Table 3a. Autocorrelation Function and Partial Auto Correlation Function for Number of crashes

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
**| . |	**| . |	1	-0.328	-0.328	6.7946	0.009
. |** |	. |*. |	2	0.227	0.133	10.090	0.006
**| . |	.*| . |	3	-0.205	-0.111	12.845	0.005
. |*. |	. |*. |	4	0.201	0.099	15.525	0.004
. | . |	. |*. |	5	-0.036	0.101	15.614	0.008
. |*. |	. | . |	6	0.082	0.045	16.080	0.013
. | . |	. | . |	7	-0.064	-0.008	16.370	0.022
.*| . |	.*| . |	8	-0.072	-0.135	16.744	0.033
. |*. |	. | . |	9	0.105	0.071	17.545	0.041
.*| . |	.*| . |	10	-0.117	-0.082	18.569	0.046
. |** |	. |*. |	11	0.215	0.150	22.077	0.024
.*| . |	. | . |	12	-0.142	0.027	23.644	0.023
. | . |	.*| . |	13	-0.007	-0.143	23.648	0.035
. | . |	. | . |	14	-0.007	0.052	23.652	0.050
. | . |	.*| . |	15	-0.046	-0.122	23.824	0.068
. | . |	. | . |	16	0.003	-0.049	23.824	0.093
. | . |	. | . |	17	-0.064	-0.037	24.175	0.115
. | . |	.*| . |	18	-0.055	-0.117	24.443	0.141
. |*. |	. |*. |	19	0.108	0.206	25.511	0.144
. | . |	. | . |	20	-0.033	-0.027	25.613	0.179
. | . |	. | . |	21	-0.027	-0.035	25.684	0.219
.*| . |	.*| . |	22	-0.074	-0.077	26.227	0.242
. | . |	.*| . |	23	0.012	-0.105	26.241	0.290
.*| . |	. | . |	24	-0.069	-0.012	26.730	0.317
. |*. |	. | . |	25	0.111	0.014	28.040	0.306
.*| . |	. | . |	26	-0.115	0.018	29.498	0.289
. | . |	. | . |	27	0.036	0.024	29.642	0.330
. | . |	. |*. |	28	0.043	0.093	29.852	0.370

Date: 05/23/23; Time: 12:04; Sample: 1960 2020; Included observations: 60

Table 3b. Autocorrelation Function and Partial Auto Correlation Function for Number of Killed

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. |*. |	. |*. |	1	0.133	0.133	1.1141	0.291
. | . |	. | . |	2	-0.045	-0.064	1.2450	0.537
.*| . |	. | . |	3	-0.068	-0.054	1.5430	0.672
. | . |	. | . |	4	-0.061	-0.049	1.7938	0.774
. | . |	. | . |	5	0.003	0.012	1.7944	0.877
.*| . |	.*| . |	6	-0.183	-0.200	4.1139	0.661
. | . |	. |*. |	7	0.060	0.114	4.3660	0.737
. |*. |	. |*. |	8	0.118	0.075	5.3553	0.719
. |*. |	. | . |	9	0.089	0.052	5.9271	0.747
. |** |	. |** |	10	0.262	0.262	11.041	0.354
. | . |	. | . |	11	0.055	0.024	11.267	0.421
. | . |	. | . |	12	-0.021	-0.029	11.301	0.503
. | . |	. | . |	13	-0.036	0.047	11.402	0.577
**| . |	**| . |	14	-0.228	-0.214	15.600	0.338
. | . |	. |*. |	15	0.067	0.134	15.976	0.384
.*| . |	.*| . |	16	-0.082	-0.079	16.543	0.416
.*| . |	.*| . |	17	-0.113	-0.189	17.649	0.411
. | . |	. | . |	18	0.022	0.010	17.692	0.476
. |*. |	. |*. |	19	0.128	0.101	19.182	0.445
. |*. |	.*| . |	20	0.121	-0.110	20.554	0.424
. | . |	. | . |	21	-0.064	0.033	20.947	0.462
.*| . |	.*| . |	22	-0.163	-0.164	23.558	0.371
. |*. |	. |*. |	23	0.076	0.120	24.143	0.396
.*| . |	. | . |	24	-0.108	-0.032	25.353	0.387
. | . |	. | . |	25	-0.013	0.024	25.371	0.442
. | . |	. | . |	26	-0.025	0.002	25.439	0.494
.*| . |	. | . |	27	-0.082	-0.031	26.202	0.507
. | . |	.*| . |	28	-0.021	-0.168	26.252	0.559

Date: 05/23/23; Time: 12:06; Sample: 1960 202
Included observations: 60

Table 4a. Automatic Model Selection for Number of Crashes

Model selected: ARIMA (1,1,0)

Parameter	***Estimate***	***Stnd. Error***	***t***	***P-value***
AR(1)	-0.332014	0.122292	-2.71492	0.008681

Backforecasting: yes

Statistic	***Estimation Period***	***Validation Period***
RMSE	3385.52	
MAE	2356.86	
MAPE	12.7402	
ME	-84.9424	
MPE	-2.36375	

RMSE: Root Mean Squared Error
MAE: Mean Absolute Error
MAPE: Mean Absolute Percentage Error
ME: Mean Error
MPE: Mean Percentage Error
Estimated white noise variance = 1.14766E7 with 59 degrees of freedom
Estimated white noise standard deviation = 3387.7

Table 4b. Automatic Model Selection for Number of Killed

Model selected: Random walk

Statistic	***Estimation Period***	***Validation Period***
RMSE	882.309	
MAE	672.65	
MAPE	11.0141	
ME	74.85	
MPE	1.83003	

RMSE: Root Mean Squared Error
MAE: Mean Absolute Error
MAPE: Mean Absolute Percentage Error
ME: Mean Error
MPE: Mean Percentage Error

Table 5a. Estimation Period for Number of Crashes

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*	*AIC*	*HQC*	*SBIC*
(A)	3559.99	2397.17	12.3268	-73.9333	-2.00659	16.355	16.355	16.355
(B)	3589.26	2395.83	12.2817	9.09495E-14	-1.54394	16.4042	16.4178	16.4388
(C)	8344.39	6931.07	41.8944	-1.19278E-12	-19.2371	18.0915	18.105	18.1261
(D)	7516.54	5864.62	33.2823	-5.06931E-13	-14.1068	17.9153	17.9424	17.9845
(E)	5757.74	4519.01	27.4919	1.55061E-12	-6.58232	17.415	17.4556	17.5188
(F)	7870.86	5974.94	31.4682	1243.0	-6.83141	18.0074	18.0345	18.0766
(G)	8545.91	6557.11	36.4217	1634.56	-8.95461	18.172	18.1991	18.2412
(H)	3450.34	2512.14	13.4148	-47.2797	-2.17925	16.3252	16.3388	16.3598
(I)	3407.76	2339.5	12.6032	-82.1169	-2.36643	16.3004	16.314	16.335
(J)	3528.88	2541.18	13.6845	42.2962	-0.149404	16.3703	16.3838	16.4049
(K)	3470.44	2388.67	12.7812	-232.796	-2.24662	16.3696	16.3968	16.4389
(L)	3711.46	2739.81	14.9754	-35.0129	-0.761209	16.4711	16.4847	16.5058
(M)	3385.52	2356.86	12.7402	-84.9424	-2.36375	16.2873	16.3009	16.3219
(N)	3359.49	2336.67	12.5399	-77.0972	-2.13927	16.3047	16.3318	16.3739
(O)	3374.62	2367.84	12.6055	-72.5683	-2.03919	16.3136	16.3408	16.3829
(P)	3436.04	2372.09	12.771	-90.7961	-2.46129	16.3169	16.3305	16.3515

Model	*RMSE*	*RUNS*	*RUNM*	*AUTO*	*MEAN*	*VAR*
(A)	3559.99	OK	OK	OK	OK	***
(B)	3589.26	OK	OK	OK	OK	***
(C)	8344.39	**	***	***	***	***
(D)	7516.54	OK	***	***	OK	***
(E)	5757.74	***	***	***	*	*
(F)	7870.86	OK	***	***	OK	***
(G)	8545.91	**	***	***	***	**
(H)	3450.34	*	*	OK	OK	***
(I)	3407.76	OK	OK	OK	OK	***
(J)	3528.88	OK	**	OK	OK	***
(K)	3470.44	OK	OK	OK	OK	***
(L)	3711.46	*	**	OK	OK	***
(M)	3385.52	OK	OK	OK	OK	***
(N)	3359.49	OK	OK	OK	OK	***
(O)	3374.62	OK	OK	OK	OK	***
(P)	3436.04	OK	OK	OK	OK	***

Models

(A) Random walk

(B) Random walk with drift = -73.9333
(C) Constant mean = 19013.7
(D) Linear trend = 25564.2 + -211.305 t
(E) Quadratic trend = 14269.1 + 864.413 t + -17.3503 t^2
(F) Exponential trend = exp(10.1386 + -0.0121281 t)
(G) S-curve trend = exp(9.77838 + -0.204758 /t)
(H) Simple moving average of 2 terms
(I) Simple exponential smoothing with alpha = 0.744
(J) Brown's linear exp. smoothing with alpha = 0.3788
(K) Holt's linear exp. smoothing with alpha = 0.7237 and beta = 0.0471
(L) Brown's quadratic exp. smoothing with alpha = 0.2615
(M) ARIMA(1,1,0)
(N) ARIMA(1,1,1)
(O) ARIMA(2,1,0)
(P) ARIMA (0,1,1)

RMSE = Root Mean Squared Error
RUNS = Test for excessive runs up and down
RUNM = Test for excessive runs above and below median
AUTO = Ljung-Box test for excessive autocorrelation
MEAN = Test for difference in mean 1st half to 2nd half
VAR = Test for difference in variance 1st half to 2nd half
OK = not significant (p >= 0.05)
* = marginally significant (0.01 < p <= 0.05)
** = significant (0.001 < p <= 0.01)
*** = highly significant (p <= 0.001)

Table 5b. Estimation Period for Number of Killed

Model	*RMSE*	*MAE*	*MAPE*	*ME*	*MPE*	*AIC*	*HQC*	*SBIC*
(A)	882.309	672.65	11.0141	74.85	1.83003	13.5651	13.5651	13.5651
(B)	886.547	658.298	10.5188	-2.72848E-13	0.222957	13.6075	13.621	13.6421
(C)	2624.18	2102.85	61.6141	4.17473E-13	-38.1385	15.7778	15.7914	15.8124
(D)	2489.01	2111.1	52.3015	-5.36751E-13	-29.0624	15.7049	15.732	15.7741
(E)	1481.9	1177.53	22.9228	-9.24404E-13	-5.36728	14.7005	14.7412	14.8043
(F)	2792.14	2322.89	45.0768	516.152	-12.5872	15.9347	15.9618	16.0039
(G)	2153.76	1710.08	34.7615	478.508	-8.62495	15.4155	15.4426	15.4847
(H)	1057.1	784.458	12.6722	109.508	2.32438	13.9593	13.9729	13.994
(I)	882.321	661.637	10.8338	73.6298	1.80014	13.5979	13.6115	13.6325
(J)	991.671	681.938	11.2906	13.4412	1.25739	13.8316	13.8451	13.8662
(K)	892.574	644.145	10.1531	-55.6043	-0.791004	13.6538	13.6809	13.723
(L)	1103.78	770.945	12.2	-2.9658	0.063503	14.0458	14.0593	14.0804
(M)	882.309	672.65	11.0141	74.85	1.83003	13.5651	13.5651	13.5651
(N)	880.364	659.002	10.7775	65.4742	1.63126	13.5935	13.607	13.6281
(O)	881.126	658.398	10.7533	64.735	1.61979	13.5952	13.6088	13.6298
(P)	890.86	664.85	10.4008	-29.3743	-0.684714	13.6172	13.6307	13.6518

Model	RMSE	RUNS	RUNM	AUTO	MEAN	VAR
(A)	882.309	OK	OK	OK	OK	OK
(B)	886.547	OK	OK	OK	OK	OK
(C)	2624.18	***	***	***	OK	***
(D)	2489.01	***	***	***	OK	*
(E)	1481.9	***	***	***	OK	OK
(F)	2792.14	***	***	***	*	OK
(G)	2153.76	***	***	***	OK	*
(H)	1057.1	**	*	OK	*	OK
(I)	882.321	OK	OK	OK	OK	OK
(J)	991.671	OK	OK	OK	OK	OK
(K)	892.574	OK	OK	OK	OK	OK
(L)	1103.78	*	OK	***	OK	OK
(M)	882.309	OK	OK	OK	OK	OK
(N)	880.364	OK	OK	OK	OK	OK
(O)	881.126	OK	OK	OK	OK	OK
(P)	890.86	OK	OK	OK	OK	OK

Models

(A) Random walk
(B) Random walk with drift = 74.85
(C) Constant mean = 6103.61
(D) Linear trend = 4547.25 + 50.205 t
(E) Quadratic trend = -93.6895 + 492.199 t + -7.12894 t^2
(F) Exponential trend = exp(8.08387 + 0.0162517 t)
(G) S-curve trend = exp(8.80048 + -2.76415 /t)
(H) Simple moving average of 2 terms
(I) Simple exponential smoothing with alpha = 0.9999
(J) Brown's linear exp. smoothing with alpha = 0.5896
(K) Holt's linear exp. smoothing with alpha = 0.9998 and beta = 0.0197
(L) Brown's quadratic exp. smoothing with alpha = 0.3199
(M) ARIMA(0,1,0)
(N) ARIMA(0,1,1)
(O) ARIMA(1,1,0)
(P) ARIMA(0,2,1)

RMSE = Root Mean Squared Error
RUNS = Test for excessive runs up and down
RUNM = Test for excessive runs above and below median
AUTO = Ljung-Box test for excessive autocorrelation
MEAN = Test for difference in mean 1st half to 2nd half
VAR = Test for difference in variance 1st half to 2nd half
OK = not significant (p >= 0.05)

* = marginally significant (0.01 < p <= 0.05)
** = significant (0.001 < p <= 0.01)
*** = highly significant (p <= 0.001)

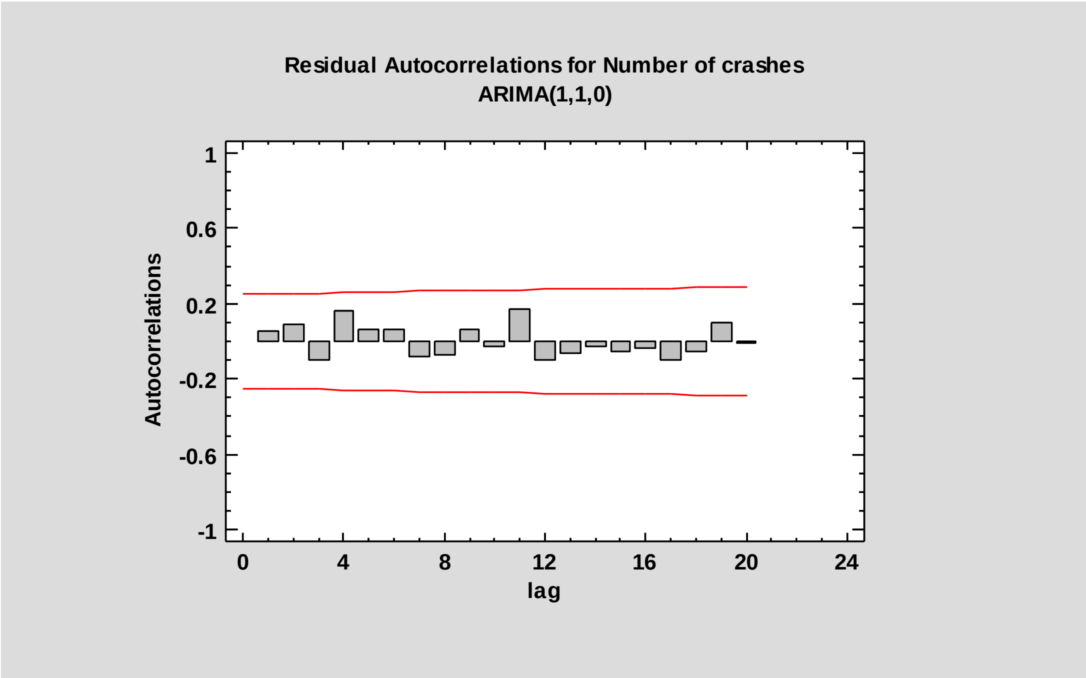

Figure 2a. Residual Autocorrelation for Adjusted Number of Crashes

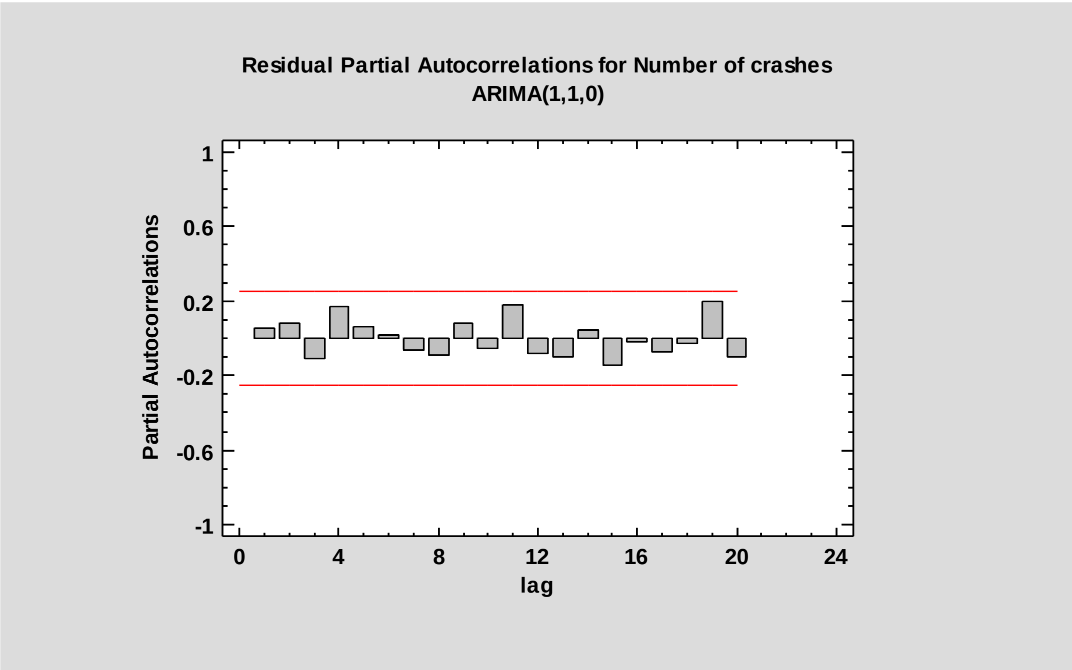

Figure 2b. Residual Partial Autocorrelations for Adjusted Number of Crashes

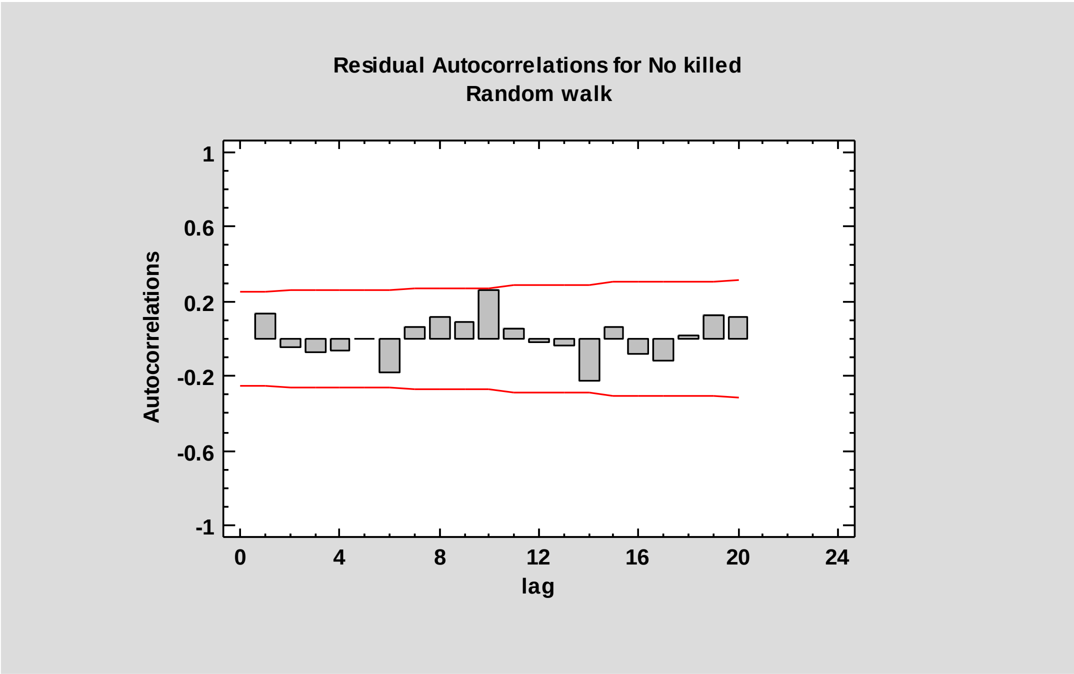

Figure 3a. Residual Autocorrelation for Adjusted Number of Killed

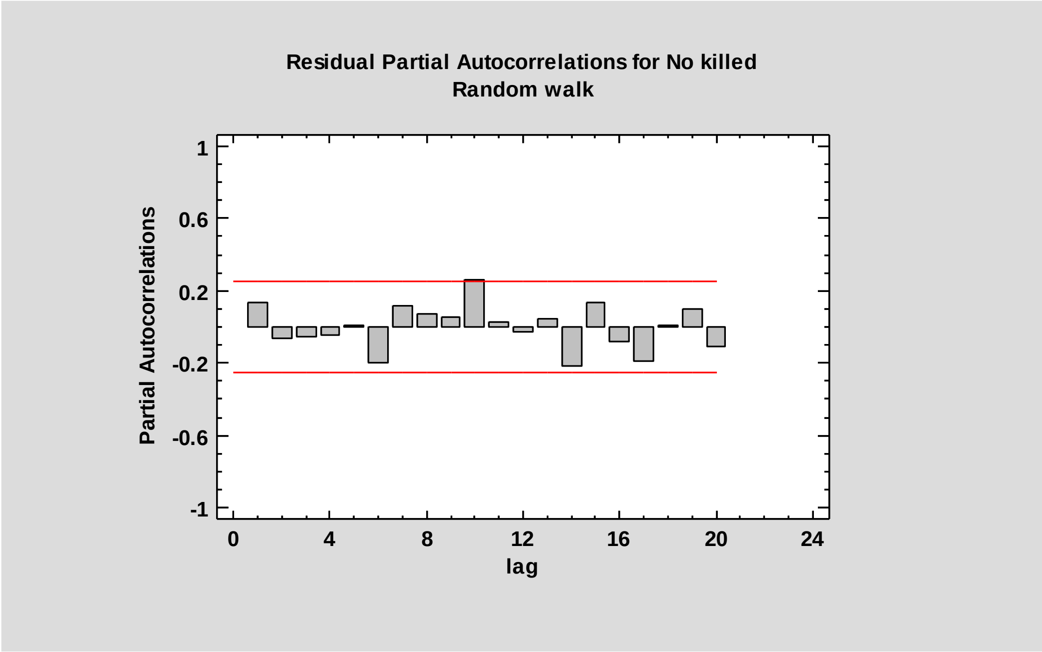

Figure 3b. Residual Partial Autocorrelations for Adjusted Number of Killed

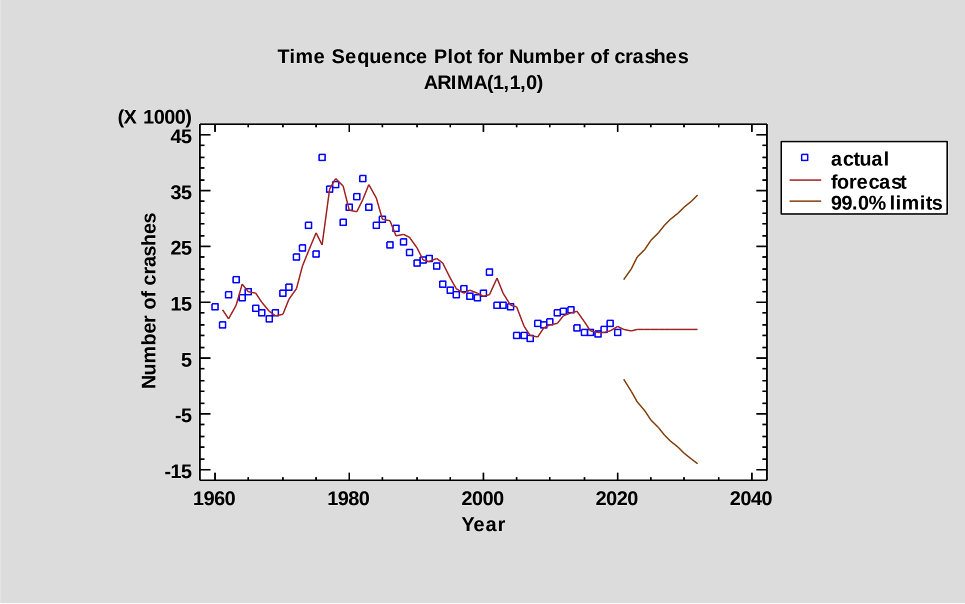

Figure 4a. Forecast Plot for Number of Crashes

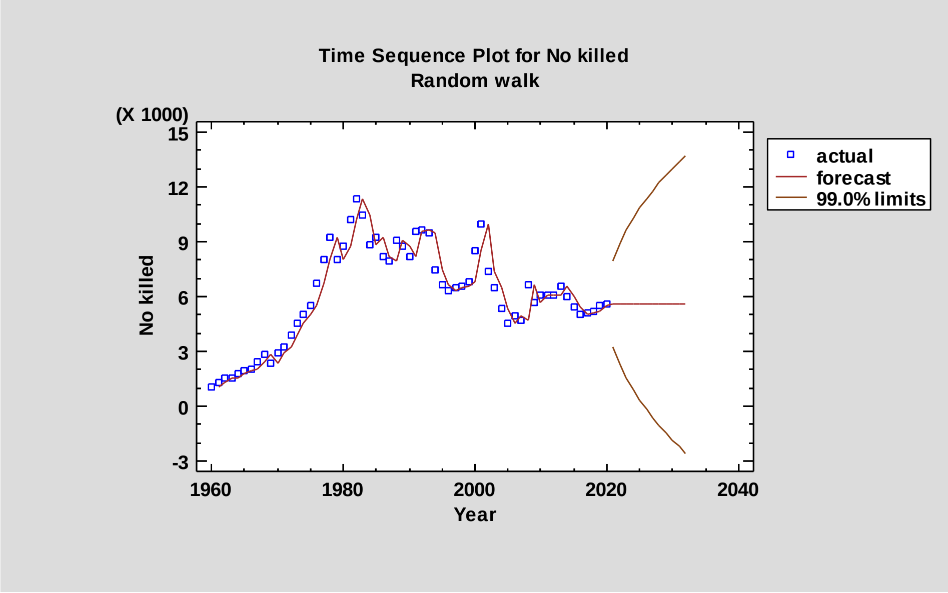

Figure 4b. Forecast Plot for Number of Killed

Results

Table 1 presents the data's mean, standard deviation, lowest, maximum, and range. The average number of road traffic crashes in Nigeria from 1960 to 2020 was 19013.7, with a standard deviation of 8344.39. The lowest number of road traffic accidents was 8477.0, while the highest number was 40881.0. The range of the data was 32404.0. Between 1960 and 2020, the average number of fatalities in road traffic crashes in Nigeria was 6103.61, with a standard deviation of 2624.18. The smallest number of fatalities was 1083.0, while the maximum was 11382.0. The range of the data was 10299.0.

Trend analysis

To examine the trend of road traffic crashes and fatalities in Nigeria, time-series data from 1960 to 2020 were plotted on a line graph, as illustrated in Figure 1.
Figure 1 shows that the number of road traffic crashes and fatalities in Nigeria increased from 1960 to 1976 before declining from 1977 to 2020. The trend line has a positive slope, showing a considerable decrease in the number of road traffic accidents and fatalities over time.

Stationarity testing

Stationarity is a fundamental assumption of time series analysis. It suggests that a series' statistical features, such as mean and variance, are constant across time. To determine stationarity, we used statistical tests such as the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. If the data was non-stationary, suitable modifications or different techniques were used to establish stationarity.
Table 2 shows the unit root test. The augmented Dickey-Fuller test was used to check for stationarity, and both data sets were determined to be stationary at first difference. The results showed that the number of collisions and fatalities were stationary at the first difference, with significant p-values. As a result, we may utilize the ACF and PACF to determine the order of our data.

Model selection

Selecting the appropriate model is crucial in time series analysis. We looked at a range of models, including autoregressive integrated moving average (ARIMA) and Holt-Winters exponential smoothing techniques. The model selection procedure involved assessing the ACF and PACF plots, evaluating the information criteria (e.g., AIC, BIC), and executing diagnostic tests to ensure the model's applicability.
Table 3(a and b) displays the ACF and PACF for the stable data. In the ARMA model, it calculates the order moving average and the autoregressive. Table 3a shows correlations between the frequency of crashes at different time lags, and further research or model development may be required to address these autocorrelation patterns. Furthermore, the Q-Stat test assesses if the associations are statistically significant. Several p-values are consistently less than 0.05, implying that the residuals are not independent. Table 3b shows that there may be correlations between the number of killed at different time lags. The Q-Stat test findings indicate that the residuals may have some autocorrelation; however, this should be interpreted in conjunction with the p-values for each lag.

Model estimation and validation

After selecting the best model, we estimated its parameters using maximum likelihood estimation or other appropriate methods. We assessed the model's performance using statistical measures like root mean squared error (RMSE), mean absolute error (MAE), and residual graphical analysis. If the model did not meet the specified requirements, we made parameter tweaks or looked into alternative models.
Table 4a illustrates the model for the number of crashes. The Autoregressive Integrated Moving Average (ARIMA) model has been used. This approach implies that the best forecast for future data is provided by a parametric model that compares the most current data value to prior data values and noise. The output summarizes the statistical significance of each term in the forecasting model. Terms with P-values less than 0.05 differ statistically substantially from zero at the 95.0% confidence level. The P-value for the AR (1) term is less than 0.05, indicating a significant difference from zero. The estimated standard deviation for the supplied white noise is 3387.71. Table 4b also displays the model for the number of people murdered. A random walk model was adopted. This model believes that the most accurate forecast for future data is provided by the last accessible data point. Table 4(a & b) highlights the present models' performance in fitting historical data. It displays RMSE, MAE, MAPE, ME, and MPE. Each statistic is based on one-ahead forecast errors, which are the discrepancies between the observed value at time t and the forecasted value at time t-1. The first three statistics assess the magnitude of the errors. A better model will result in lower value. The final two statistics quantify bias. A better model will return a value close to zero.

Model comparison for number of crashes

Table 5a compares the results of fitting multiple models to data on the frequency of crashes. Model M (ARIMA (1,1,0), which was used to generate the forecasts, has the lowest Akaike Information Criterion (AIC) value. Table 5a also shows the results of five residual tests performed to determine whether each model is suitable for the data set. An OK means the model passed the test. "*" means that it fails with a 95% confidence level. "**" implies that it fails with a 99% confidence level. "***" implies that it fails with a 99.9% confidence level. Model M, the current model under consideration, has passed four tests.

Model comparison for number of killed

Table 5b compares the results of fitting several models to data on the number of deaths. Model A (Random Walk Model) has the lowest Akaike Information Criterion (AIC) value and was utilized to generate the projections. Table 5b further highlights the results of five residual-based tests to determine whether each model is adequate for the data. An OK indicates that the model passed the test. "*" indicates that it fails at 95% confidence level. "**" indicates that it fails at 99% confidence level. "***" indicates that it fails with a 99.9% confidence level. It is worth noting that the currently selected model, model A, passes five tests. Since no tests are statistically significant at the 95% or higher confidence level, the existing model is most likely suitable for the data.
Figures 2(a & b) and 3(a & b) show the estimated autocorrelations and partial autocorrelations between the residuals at different delays. The lag k autocorrelation coefficient computes the correlation between the residuals at time t and time t-k. Also included are 95.0% probability bounds around 0. If the probability limits at a specific lag do not include the predicted coefficient, the lag has a statistically significant association at the 95.0% confidence level. The lack of notable spikes in ACF and PACF plots for residuals in Figures 2(a &b) and 3(a &b) is a positive result, indicating that the time series model is well-specified and has effectively captured the data's temporal dynamics. This is critical for developing accurate forecasts and reaching valid findings from the analysis.
Figure 4 (a & b) depicts the expected number of crashes and fatalities. The graphs also provide 99.0% prediction limitations for the forecast. These boundaries indicate where the true number of crashes and fatalities are most likely to occur in the future with 99.0% confidence.

Discussion

The study investigated and forecasted the trend of road traffic crashes and fatalities in Nigeria, developing an ARIMA and Random walk model from 1960 to 2020. The ARIMA model created in this study was an ARIMA(1,1,0) model, indicating that the time-series data contained an autoregressive (AR) component of lag 1, a differentiating component of order 1, and a moving average (MA) component of lag 0. The ARIMA (1,1,0) model had AIC and BIC values of 16.2873 and 16.3219, indicating a satisfactory fit to the data. To validate the ARIMA model, the fitted values were compared to the real time-series data values. The model performed well on the data, with a Mean Absolute Percentage Error (MAPE) of 12.7% for the number of road traffic crashes.
Furthermore, the model designed for the number of killed was a random walk model. The Random walk model had AIC and BIC values of 13.5651 and 13.5651, respectively, indicating a satisfactory fit to the data. To validate the Random walk model, its fitted values were compared to the actual values of the time-series data. The model fit the data well, with a mean absolute percentage error (MAPE) of 11.01% for the number of people died in road crashes.
The ARIMA and Random Walk models were then used to anticipate the pattern of road traffic crashes and fatalities in Nigeria over the following twelve years (2021-2032). Figure 4 (a&b) shows that the number of road traffic crashes and fatalities in Nigeria is expected to decrease between 2021 and 2032.

Conclusion

The analysis discovered that the trend of road traffic crashes and fatalities in Nigeria grew for several years before declining. The rise in the tendency can be ascribed to a variety of issues, including bad road infrastructure, a lack of road safety education, irresponsible driving, over speeding, and inadequate law enforcement. The ARIMA and Random walk models built in this study revealed that the trend is projected to reverse from 2021 to 2032. This study has shed light on the pattern of road traffic crashes and fatalities in Nigeria. The findings of this study indicate that measures targeted at lowering road traffic collisions and fatalities in Nigeria have been implemented and are projected to provide favorable outcomes in the coming years.

References

1. World Health Organization. Global status report on road safety 2018. Available from: https://www.who.int/violence_injury_prevention/road_safety_status/2018/en.

2. Sadeghi-Bazargani, H., Ayubi, E., Azami-Aghdash, S., Abedi, L., Zemestani, A., Amanati, L. Epidemiological patterns of road traffic crashes during the last two decades in Iran: A review of the literature from 1996 to 2014. Arch Trauma Res 2016;5(3):e32985.

3. Ma, W., Wang, Y., Bai, L., Shi, J. Road traffic injury in China: A review of national data sources. Traffic Injur Prevent 2012;13(6):544-549.

4. Bener, A., Rahman, Y.S., Mitra, B., Al-Thani, M. Incidence and severity of head and neck injuries in victims of road traffic crashes: In an economically developed country. Int Emerg Nurs 2009;17(1):52-59.

5. Kumar, A., Kumar, S., Kumar, S. An exploration of characteristics and time series forecast of fatal road crashes in Manipal, India. Sustainability 2022;14(5):2851.

6. Dharmaratne, S.D., Jayatilleke, A.U., Jayatilleke, A.C. Road traffic crashes, injury and fatality trends in Sri Lanka: 1938–2013. Bull WHO 2015;93(6):403-411.

7. Ning, B., Li, Y., Li, Z. Analyzing the impact of traffic violation behaviors on traffic fatal crashes using multilevel models on expressways. Math Problem Engineer 2022:9917877.

8. Blankson, A.K., Lartey, M.Y. Road traffic accidents in Ghana: Contributing factors and economic consequences. Ghana Med J 2020;54(3):161-168.

9. Adams, S.O., Haruna, Y.U., Mohammed, T. Cluster Analysis of HIV/AIDs Incidence in Sub-Saharan Africa (1990 – 2018). Int J Epidemiol Health Sci 2023;4: e54.

10. Blankson, A.K., Addo, V.N., Agyemang, W. Epidemiology of injuries presenting to the accident center of Korle-Bu Teaching Hospital, Ghana. BMC Emerg Med 2019;19(1):1-8.

11. Tanimu M, Yahaya H.U., Samuel O.A. Modeling the Volatility for Some Selected Beverages Stock Returns in Nigeria (2012-2021): A Garch Model Approach. Matrix Sci Math 2022;6(2):41-51.

12. Box, G.E.P, Jenkins, G.M., "Time series analysis: "Forecasting and control," Holden-Day, San Francisco,1976.

13. Yahaya H.U. Tanimu, M. Analysis of Mortality Rate in Nigeria. Int J Sci Techno 2016;4(11):80-89.

14. Akaike, H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974; 19(6):716-723.

15. John, E.I. Effect of Macroeconomic Variables on Stock Market Performance in Nigeria. J Economic Manage Trade 2019;22(6):1-14.

16. Mann, H. B. Nonparametric tests against trend. Econometrica 1945;13:245–259.

17. Ke, Z., Zhang, Z. Testing Autocorrelation and Partial Autocorrelation: Asymptotic Methods versus Resampling Techniques. Br J Math Statistic Psychol 2018;71(1):96-116.

18. Yahaya, H.U., Olanrenwaju, S.O., Mohammed, T. Testing Linkages amongst Emerging African Markets. Int J Sci Technol 2017;5(6):1-11.