African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)
www.abjournals.org

VECTOR AUTOREGRESSION MODELING OF MALARIA INCIDENCE AND MORTALITY RATES IN MIGORI COUNTY, KENYA: A TIME SERIES ANALYSIS INCORPORATING EXOGENOUS INFLUENCES

Chacha Paul Jackson[1*], Argwings Otieno (Ph.D.)[2], and Julius Koech (Ph.D.)[3]

[1-3]Department of Mathematics and Computer Science, University of Eldoret, Kenya.

Emails: [1]mahirijackson@gmail.com; [2]ranyot2001@yahoo.com; [3]jkoech@uoeld.ac.ke

[*]Corresponding Author's Email: mahirijackson@gmail.com

Cite this article:

Jackson, C. P., Otieno, A., Koech, J. (2025), Vector Autoregression Modeling of Malaria Incidence and Mortality Rates in Migori County, Kenya: A Time Series Analysis Incorporating Exogenous Influences. African Journal of Mathematics and Statistics Studies 8(3), 31-46. DOI: 10.52589/AJMSS_MK6PPMDN

Manuscript History

Received: 2 Feb 2025

Accepted: 7 Mar 2025

Published: 25 Jul 2025

ABSTRACT: *Malaria remains hyperendemic in Kenya's Lake Victoria basin despite scalable interventions. It is a pressing public health challenge in Migori county that reported a 27% mortality rate in 2020 in children aged 6-59 months, far exceeding national levels. Reports indicate different contributing factors to malaria dynamics in Migori County, including marginal insecticide-treated net (ITN) use, ITN access, effective anti-malaria treatment, and prevalence of malaria infection. The present study seeks to elucidate the temporal interaction between malaria incidence and mortality by employing a range of time series analyses, incorporating exogenous influences by applying classical vector autoregressive (VAR) model to capture lagged dependencies. Further, the study invoked a Bayesian VAR (BVAR) after incorporating exogenous variables for parameter estimating, utilizing Monte Carlo simulations and Gibbs sampling. For model adequacy and forecast accuracy, the analysis made use of Ljung-Box test, partial autocorrelation function, autocorrelation function (ACF), and normality tests among other diagnostic tests. The hierarchical Bayesian vector autoregressive model (BVARX) incorporates monthly incidence and mortality rates (2014-2024,* $n = 120$*) as the endogenous variables. The exogenous variables comprised ITN access and use, treatment efficacy, and infection prevalence. Ward-level heterogeneity summed the spatial random effects. Hamilton Monte Carlo model estimation with convergence assessed using* $\hat{R} < 1.01$ *Counterfactual simulations quantified intervention impacts. ITN use reduced incidence (β = −1.43, 95% CrI: −2.21, −0.65) but access increased mortality (β = 1.81, CrI: 0.32, 3.30), suggesting behavioral misuse. VARX outperformed VAR (WAIC: 412 vs. 587), yet residual spatial autocorrelation (Moran's I = 0.34, *p* = 0.01) indicated unobserved confounders. BVARX forecasts predicted 22% (CrI: 18–27%) higher incidence by 2025 under current interventions. The regression analysis identified that higher ITN use is significantly associated with reductions in both malaria mortality and incidence. While ITNs and treatments show efficacy, their benefits are eroded by suboptimal utilization and ecological feedback. The study recommended the use of ward-level VARX outputs for geospatial targeting of ITN campaigns as well as integrated resistance monitoring through adaptive Bayesian frameworks.*

KEYWORDS: Malaria Elimination, Causal Time Series, Hierarchical VAR, VARX, BVAR, Intervention Optimization.

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)

www.abjournals.org

INTRODUCTION

In sub-Saharan Africa regions where climatic, socioeconomic, and intervention factors interact to produce highly variable transmission patterns, malaria remains a leading cause of morbidity and mortality. Despite extensive implementation of control measures, significant heterogeneity persists across counties in Kenya. Migori County in Kenya has a high burden of malaria owing to its location near Lake Victoria, where climatic conditions characterized by tropical climatic conditions favor the breeding of Anopheles mosquitoes (Guissou et al., 2023; Okiro et al., 2010; Ototo et al., 2022). Vulnerable groups, including children, suffer the most, informing the need for robust forecasting models that, besides accounting for inherent temporal dynamics, integrate the effects of fundamental public health interventions.

Conventional epidemiological models fail to capture the bidirectional feedback that may occur between incidence and mortality series due to reliance on univariate techniques, such as the ARIMA or Holt's model (Savi, 2023). To create a framework that models the intricate relationship among multiple time series, advanced multivariate techniques, such as vector autoregression (VAR) models, may offer significant insights to guide public health interventions. The preference for VAR models stems from the flexibility of variables to function on their past values, as well as historical data of other variables in the system. VAR models have performed well and exhibited good fit when applied by different researchers to forecast infectious diseases and macroeconomic indicators (Oyegoke et al., 2023). However, their application in malaria epidemiology, particularly in the context of incorporating exogenous intervention variables, remains limited as it is still in its infancy.

Our study employs a sequential modeling approach to address this gap. In doing so, we first estimate a VAR model to explore the underlying correlations between malaria incidence and mortality rate in Migori County. The VAR model is then progressively augmented with exogenous variables (ITN use, ITN access, effective anti-malaria treatment, and malaria infection prevalence) to generate a VARX model. Subsequently, we will deploy a Bayesian VAR (BVAR) model to leverage prior information and formulate posterior inference through Monte Carlo simulations. Finally, to ensure that the intrinsic assumptions hold, diagnostic tests including checks for stationarity, serial correlation, and the normality of residuals will be performed.

The study will seek to satisfy three objectives. First, we purposely elucidate the complex association between malaria incidence and mortality, flagging any lagged feedback mechanisms. Second, we incorporate exogenous variables into the modeling framework to evaluate the direct influence of intervention measures on these series. Finally, we seek to assess the robustness and forecasting performance of these models by comparing the output from classical and Bayesian formulations of the VAR model in respect to capturing the complex epidemiological dynamics of malaria in Migori County.

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)

www.abjournals.org

LITERATURE REVIEW

With researchers employing different statistical and machine learning models to understand and predict disease outbreaks, the application of time series analysis in epidemiology has grown significantly in the recent past. The most widely used time series models for forecasting disease incidence are the Autoregressive Integrated Moving Average (ARIMA), as used by Achieng et al. (2020) to identify the determinants influencing the incidence of malaria in Kisumu County, with good performance and fit in studying malaria, dengue, and influenza. However, due to the univariate form of ARIMA models, i.e., they analyze a single time series at a time, their capacity is limited, failing to capture the convoluted dynamics of multiple epidemiological factors, an aspect that the present study aims to bridge.

Evidence points to the efficacy of multivariate time series approaches like VAR in public health research that have gained traction in addressing univariate model limitations. VAR models are particularly useful when multiple variables are simultaneously influenced by each other or when there is no solid theoretical basis to distinguish between endogenous and exogenous variables. Badaoui et al. (2023) and Bayarbat and Li (2020) successfully applied VAR models in various health contexts, such as forecasting hospital admissions, analyzing the relationship between air pollution and respiratory diseases, and even exploring the dynamics of other infectious diseases. Different other studies have explored the relationship between climate variables (temperature and rainfall) and malaria incidence using VAR models, highlighting the potential for improved predictive accuracy when multiple interacting factors are considered.

Several studies have explored the dynamic relationships between incidence, mortality, and various environmental or socioeconomic determinants, specifically in the context of malaria. Darkoh et al. (2017), Diao et al. (2023), Eunice et al. (2017), Hou et al. (2023), Karlsson et al. (2023), Shah et al. (2024), Warsono et al. (2019), and Zheng (2024) are some of the pool of researchers that have incorporated VAR models in the study of malaria incidence and mortality rates, underscoring the necessity of robust statistical methods that can account for the lagged effects and feedback loops inherent in the transmission of malaria. Our exploration builds on this foundation of VAR frameworks by specifically focusing on Migori County (high-burden region), Kenya, in establishing the underlying association between malaria incidence and mortality rates. Subsequently, the study will contribute to existing literature in a dimension that enhances the understanding of how changes in one indicator might predict future changes in the other, offering a data-driven approach to enhance disease surveillance and intervention strategies.

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)

www.abjournals.org

METHODOLOGY

Area of Study and Data

The study identified Migori County, Kenya, as one of the regions with malaria endemic and resident to an elevated proportion of malaria incidences and deaths. Secondary data were extracted from the Kenya National Health Management databases. The time series records were extracted through web scraping focusing on incidences of malaria and mortality, as well as intervention related factors such as malaria infection prevalence, indicators of effective anti-malaria treatment, ITN access, and proportion of ITN use. An analysis of summary and descriptive statistics indicated that rates of malaria mortality ranged between 6.85 and 19.37 ($\mu = 13.07$), whereas malaria incidences ranged between 87.14 and 199.58 ($\mu = 142.57$), with ITN use averaging about 59.64% and ITN access approximately 61.68%; intervention metrics exhibited a similar variation.

Data Sources, Preprocessing and Stationarity Testing

Data on malaria outcomes was scrapped from the Kenya Health Information System for the period between 2014 and 2024. The Demographic Health Surveys (DHS) databases and report programs provided data on different malaria intervention initiatives in Migori County. The time series data were subjected to validation and authentication steps, which included addressing absent values, remedying outliers and ensuring consistency. The major assumption for stationarity testing before VAR modeling was constant statistical properties, implying that parameter estimates (mean, variance, autocorrelation, etc.) do not change over the study period, for the avoidance of spurious regressions. For both mortality series and incidences of malaria, Augmented Dickey-Fuller (AUD) and Phillips-Perron (PP) tests were undertaken to affirm stationarity. Incidences of non-stationarity were addressed through order of differencing by applying repeated tests, which achieved stationarity.

Regression Modeling

A simple regression framework was constructed to evaluate the independent effects of intervention variables prior to the multivariate time series on malaria outcomes. The model was defined as follows:

$$Y_{it} = \alpha_{it} + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + e_{it}$$

where Y_{it} denoted either the mortality or incidence rate, and X_{1it} through X_{4it} corresponded to ITN use, ITN access, anti-malaria treatment, and malaria infection prevalence, respectively. The coefficients were derived using Ordinary Least Squares (OLS) estimation as follows:

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\left(Y_{it} - \underline{Y_t}\right)\left(X_{it} - \underline{X_t}\right)}{\sum_{i=1}^{n}\left(X_{it} - \underline{X_t}\right)^2}$$

$$\hat{\alpha} = \underline{Y_t} - \frac{\sum_{i=1}^{n}\left(Y_{it} - \underline{Y_t}\right)\left(X_{it} - \underline{X_t}\right)}{\sum_{i=1}^{n}\left(X_{it} - \underline{X_t}\right)^2}\underline{X_t}$$

for which $\underline{Y_t} = \frac{1}{n}\sum_{i=1}^{n} Y_{it}$ and $\underline{X_t} = \frac{1}{n}\sum_{i=1}^{n} X_{it}$.

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)

www.abjournals.org

Vector Autoregression Modeling

The study utilized a vector autoregression (VAR) model to incorporate the dynamic feedback between malaria incidence and mortality as follows: Let $Y_t = [y_{1,t}\ y_{2,t}]$ be a vector that includes the malaria incidence (Series 1) and mortality (Series 2) rates observed at time t. The VAR p model was specified as

$$Y_t = \mu + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t, \quad t = 1,2,\ldots,T$$

where μ is a 2 dimensional intercept vector, A_i for $i = 1,\ldots,p$ are 2×2 time-invariant coefficient matrices that evaluate the long-run co-movement between time series variables at time t and those at time $t - i$, and ε_t is a 2×1 zero-mean white noise vector process (serially independent) with a time invariant variance-covariance matrix Σ (Chang & Shi, 2024).

VAR with Exogenous Variables (VARX)

The general form of the vector autoregressive model with exogenous variables (VARX) with the order of the endogenous variables p and the order of the exogenous variables q, VARX(p, q), is expressed as follows:

$$z_t = \mu + \sum_{i=1}^{p} \phi_i z_{t-i} + \sum_{j=0}^{q} \theta_j x_{t-j} + e_t,$$

where z_t represents the n-dimensional vector containing the endogenous variables at time t, $\Psi_i,\ i = 1,2,\ldots,p$ express the $(n \times n)$ coefficient matrices of z_{t-i}, x_t represents the r-dimensional vector containing the exogenous variables at time t, $\theta_j, j = 0,1,\ldots,q$ express the $(n \times r)$ coefficient matrices of x_{t-j}, and e_t represents the n-dimensional random error vector. Equation (1) can also be reformulated as follows:

$$\phi(B_z) z_t = \mu + \theta(B_x) x_t + e_t,$$

where $\phi(B_z) = I_n - \sum_{i=1}^{p} \phi_i B_z^i$, $\theta(B_x) = \sum_{j=0}^{q} \theta_j B_x^j$, $B_z^i z_t = z_{t-i}$, and $B_x^j x_t = x_{t-j}$.

There are three assumptions of the VARX model presented as follows:

i. The VARX(p, q) model is stationary if the roots of the determinants of $\phi(v)$ and $\theta(w)$ are outside the unit circle.

ii. e_t is independent and identically multivariate normally distributed with zero-vector mean and a constant covariance matrix Σ. Σ is a positive semi-definite matrix expressed as: $\Sigma = [Cov(e_{k1,t}, e_{k2,t})], where k_1, k_2 = 1,2,\ldots,n.$

iii. e_t and x_t are independent.

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)
www.abjournals.org

Bayesian VAR with Exogenous Variables (BVARX)

The Bayesian VAR model was defined with three different priors: independent Normal-Wishart prior, the Minnesota prior, and the SSVS prior. The VAR model was formulated in matrix form as follows:

$$Y = X\Phi + \epsilon$$

where the $T \times n$ matrix Y is defined as $Y = (y_1, \dots, y_T)'$, the $T \times (1 + np)$ matrix X is defined as $X = (x_1, \dots, x_T)'$; the $(1 + np) \times 1$ vector is defined as $x_t = \left(1, y'_{t-1}, \dots, y'_{t-p}\right)'$; the $(1 + np) \times n$ matrix Φ is defined as $\Phi = \left(\mu', \Theta'_1, \dots, \Theta'_p\right)'$; and ϵ is a $T \times n$ matrix with $\epsilon = (\epsilon_1, \dots, \epsilon_T)'$ (Hou et al., 2023).

Based on the VAR model, the study described the independent Normal-Wishart prior used in the study:

$$vec(\Phi) \sim MN(vec(\Phi_0), V_0)\ \Sigma \sim IW(\Sigma_0, \nu_0)$$

where MN refers to a multivariate normal with mean $vec(\Phi_0)$ and covariance matrix V_0, and IW refers to an inverted Wishart distribution with parameters Σ_0 and degrees of freedom ν_0. Unlike the natural conjugate priors, the prior for Φ and Σ are independently specified (Sugita, 2022).

With the joint prior and the likelihood, the conditional posterior densities of $vec(\Phi)$ and Σ are derived as follows:

$$vec(\Phi)|\Sigma, Y \sim MN(vec(\Phi^*), V^*)\ \Sigma|\Phi, Y \sim IW(\Sigma^*, \nu^*)$$

where $V^* = (V_0^{-1} + \Sigma \otimes X'X)^{-1}$ and

$$vec(\Phi^*) = V^*[V_0^{-1}vec(\Phi_0) + (\Sigma \otimes I_\kappa)^{-1}vec(X'Y)],$$

$\Sigma^* = (Y - X\Phi)'(Y - X\Phi) + \Sigma_0$, and $\nu^* = T + \nu_0$. Given these conditional posterior specifications, the Gibbs sampler generates sample draws.

Note that, with zero prior mean $\Phi_0 = 0$ and large prior variance V_0, the posterior mean for Φ is almost identical to the maximum likelihood estimator (Sukono et al., 2023). In this study, the hyperparameters were set at $vec(\Phi_0) = 0$ and $V_0 = 100$, $\Sigma_0 = 0.1I$, and $\nu_0 = 5$.

Forecasting

For a VARX (p,q) model with known parameters ϕ_i for $i = 1,2, \dots, p$ and ψ_j for $j = 1,2, \dots, q$, the best predictor for Γ_{T+1} (one-step forecast) based on data available at time T was:

$$\hat{\Gamma}_{T+1|T} = c + \Phi_1\hat{\Gamma}_T + \Phi_2\hat{\Gamma}_{T-1} + \cdots + \Phi_p\hat{\Gamma}_{T-p+1} + \Psi_1\hat{X}_T + \Psi_2\hat{X}_{T-1} + \cdots + \Psi_q\hat{X}_{T-q+1}.$$

For longer forecasts, such as h-step ahead forecasts, the chain rule of forecasting was used:

$$\hat{\Gamma}_{T+h|T} = c + \Phi_1\hat{\Gamma}_{T+h-1} + \Phi_2\hat{\Gamma}_{T+h-2} + \cdots + \Phi_p\hat{\Gamma}_{T+h-p} + \Psi_1\hat{X}_{T+h-1} + \Psi_2\hat{X}_{T+h-2} + \cdots + \Psi_q\hat{X}_{T+h-q}.$$

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)
www.abjournals.org

RESULTS AND FINDINGS

Descriptive Analysis

Table 1: Malaria Mortality and Incidence Rates

	Malaria Mortality Rate	Malaria Incidence Rate	Insect Treated Net Use	Insect Treated Net Access	Anti-Malaria Effective Treatment	Malaria Infection Prevalence
Min.	6.85	87.14	41.11	42.48	47.35	4.778
1st Qu.	11.23	106.15	53.71	55.97	51.71	5.979
Median	12.82	126.64	61.96	62.25	54.55	7.723
Mean.	13.07	142.57	59.64	61.68	54.24	8.941
3rd Qu.	14.79	187.31	68.53	71.96	57.17	12.347
Max.	19.37	199.58	75.81	79.86	58.56	13.449

As *Table 1* above summarizes, the exploratory data analysis revealed that mortality rates due to malaria in Migori County ranged from 6.85 to 19.37, with an average of 13.07. The rate of malaria incidence had a mean of 142.57, varying from 87.14 to 199.58. The prevalence of malaria infection ranged from 4.778 to 13.449 while intervention variables showed substantial variation. For instance, ITN use and ITN access had medians of 61.96% and 62.25% respectively. While the incidence and mortality data exhibited a normal distribution, histogram plots indicated that the distribution of infection prevalence was slightly skewed as displayed below:

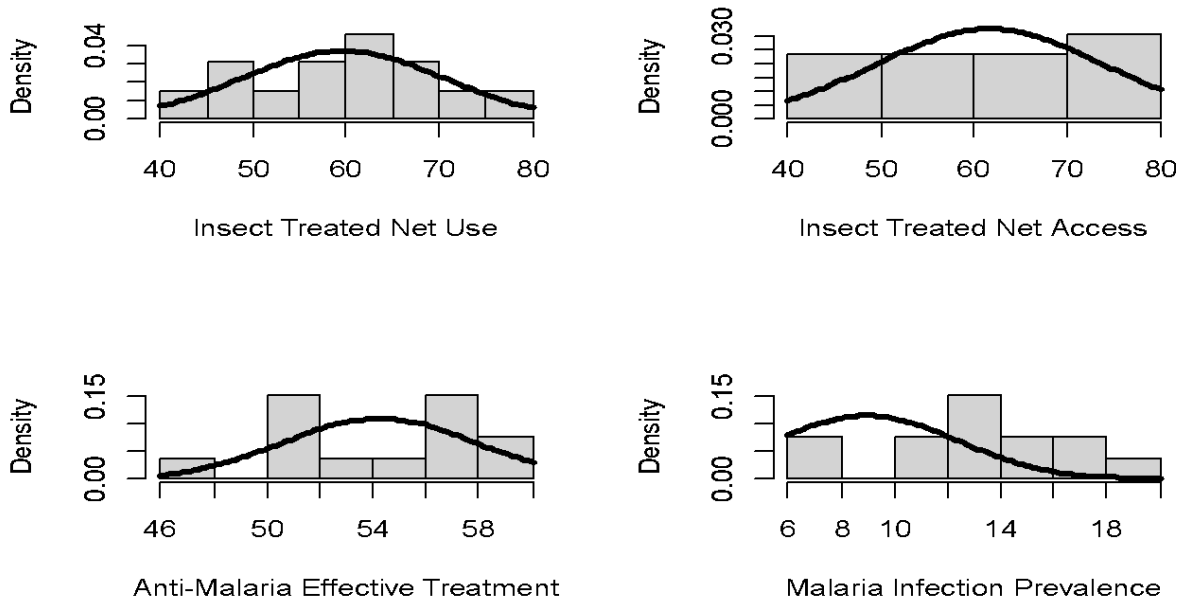

Figure 1: Descriptive Histograms

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)

www.abjournals.org

Regression Analysis

The relationship between intervention measures and malaria outcomes was modeled using base regression, providing crucial insights. With mortality as the response variable, the model exhibited a statistically significant negative effect of the coefficient on ITN use ($\beta = -2.0500, p \leq 0.0175$) and a significant positive effect from ITN access ($\beta = 1.8128, p = 0.0169$).

Table 2: Regression Model Results for Mortality Rate Response

	Estimate	Std Error	t-value	Pr(>|t|)
Intercept	-24.5374	29.4675	-0.833	0.4292
Insect Treated Net Use	-2.0500	0.6870	-2.984	0.0175**
Insect Treated Net Access	1.8128	0.6033	3.005	0.0169**
Anti-Malaria Effective Treatment	0.7916	0.4929	1.606	0.1470
Malaria Infection Prevalence	0.5728	0.5022	1.140	0.2871

Contrarily, despite being positive, the coefficients for anti-malaria effective treatment and infection prevalence lacked statistical significance. The incidence model exhibited similar trends, with ITN use bearing a close association with cases of malaria reduction ($\beta = -1.4282, p = 0.00545$). ITN access similarly had a positive association with malaria reduction cases ($\beta = 1.3114, p = 0.00427$). The findings demonstrated a strong influence of malaria infection prevalence on the incidence rate with $\beta = 13.2330$ and $p < 0.0001$ as shown in *Table 3* below.

Table 3: Regression Model Results for Malaria Incidence Rate Response

	Estimate	Std Error	t-value	Pr(>|t|)
Intercept	-1.8888	16.2433	-0.116	0.91030
Insect Treated Net Use	-1.4282	0.3787	-3.772	0.00545 **
Insect Treated Net Access	1.3114	0.3326	3.944	0.00427 **
Anti-Malaria Effective Treatment	0.5610	0.2717	2.065	0.07281
Malaria Infection Prevalence	13.2330	0.2769	47.798	<0.0001 ***

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)

www.abjournals.org

VAR and VARX Model Estimates

The dynamic interplay between malaria incidence and mortality was analyzed using the VAR model without exogenous variables. The incidence series produced a significant positive coefficient $(0.8537, p < 0.001)$ in one specification, whereas the mortality series yielded a negative coefficient $(-3.2028, p = 0.081)$, an indication that historic incidence values strongly forecast current incidence levels, while past mortality influences future mortality with marginal significance.

Table 4: Vector Autoregression with Exogenous Variables

	Incidence Series	Mortality Series
Series 1.11	-0.015748690	-0.4767374
Series 2.11	0.450001249	2.6487354
Series 1.12	0.001823247	-0.1241577
Series 2.12	0.114397189	-1.2180552
Constant	39.872684844	492.2120343
Insect Treated Net Use	-0.868202738	-1.2411085
Insect Treated Net Access	0.717548400	1.4198022
Anti-Malaria Effective Treatment	-0.222567827	-7.9697544
Malaria Infection Prevalence	13.204835786	1.6357221

The findings revealed deeper insights upon the introduction of exogenous variables via the VARX analysis. The ITNs use continued to impart a significant negative influence of -0.8682 in the incidences' equations, even when controlling for exogenous factors. On the other hand, ITN access and the prevalence of malaria infection had a positive correlation. ITN use retained a negative effect (-1.2411) for mortality equations while treatment of malaria yielded material negative coefficient (-7.9698) and ITN access rendered a positive effect (1.4198). Significant residual autocorrelation remained $(p < 0.005\ for\ both\ series)$ although the VARX model improved the overall fir compared to the baseline VAR model.

Bayesian VAR Findings

Table 5: BVAR Median Values Coefficients

	Var 1	Var 2
Constant	36.684	3.530
var1-lag1	0.880	0.002
var2-lag1	-1.877	0.706

The stability and uncertainty of coefficient estimates was provided by the BVAR analysis. After tuning the hyperparameter lambda $(\lambda = 0.15497)$, the Gibbs sampler (with a 97.4% acceptance rate) produced posterior median values that reinforced the primary relationships

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)
www.abjournals.org

observed in the classical models. The cross-variable effect of lagged mortality on current incidence was negative ($median = -1.877$) as the lagged incidence variable had a median coefficient of 0.880. The second series depicted better stability ($variance = 10.482$) while residual variability in the first series was high ($variance = 183.153$) as indicated by the variance-covariance matrix estimates. Trace plots and density plots confirmed that the chains converged satisfactorily, lending support to the robustness of the BVAR estimates.

Table 6: Median Values for Variance-Covariance Values from a BVAR

	Var 1	Var 2
Var 1	183.153	13.259
Var 2	13.259	10.482

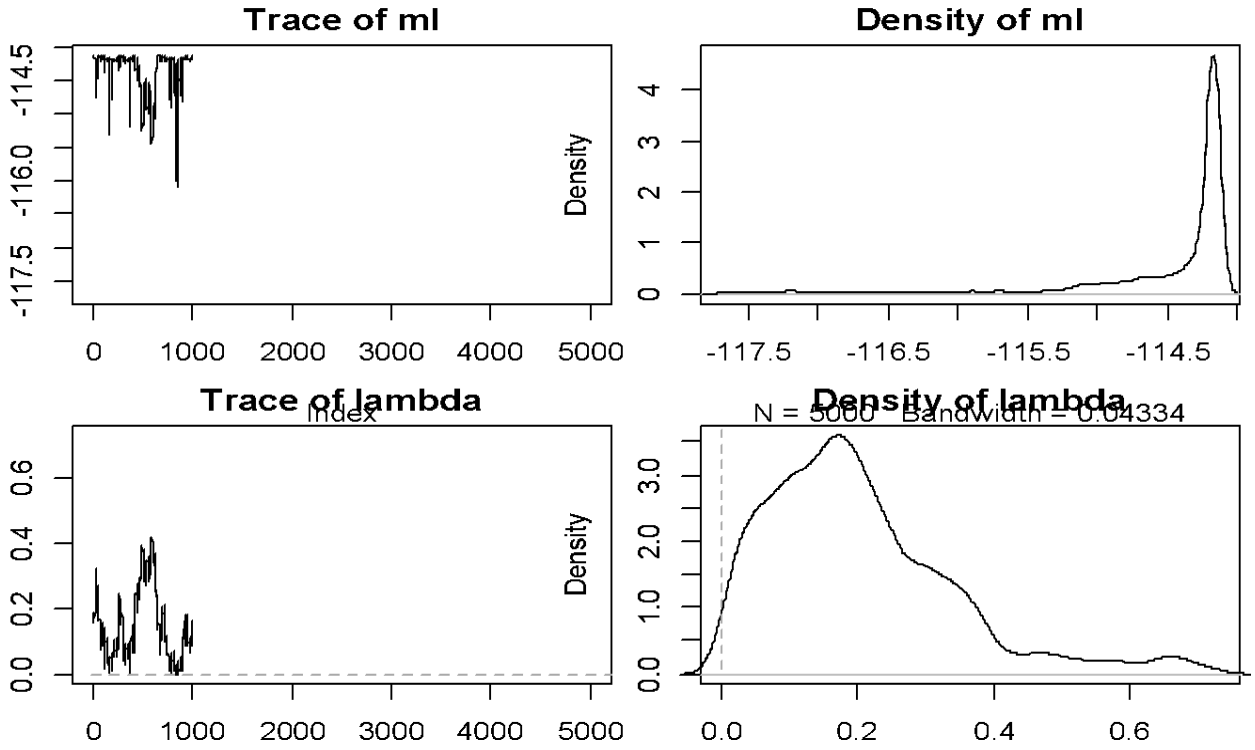

Figure 2: Trace and Density Plots

Residual Diagnostics and Model Fit

Regression models yielded residuals that were approximately normally distributed based on residual analyses across models, evidenced by Shapiro-Wilk and Lilliefors tests (Figure 3). However, in the incidence series, the VAR and VARX models displayed contained notable autocorrelation in the residuals. The Ljung–Box test yielded $p = 0.03465$ for the VAR model residuals of Series 1 and $p = 0.00316$ for the VARX model residuals of incidence. Further model refinement is needed to fully eradicate serial correlation in as much as the inclusion of exogenous variables reduced the range of the residuals significantly, i.e., the VARX residuals for Series 1 ranged from –0.77912 to 0.89667. Validation of the regression model through ANOVA demonstrated significant drivers of both mortality and incidence, particularly for ITN

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 3, 2025 (pp. 31-46)
www.abjournals.org

use and anti-malaria treatment predictors, with Pillai trace values exceeding 0.96 and significant F-statistic with 95% confidence of not committing a Type I error.

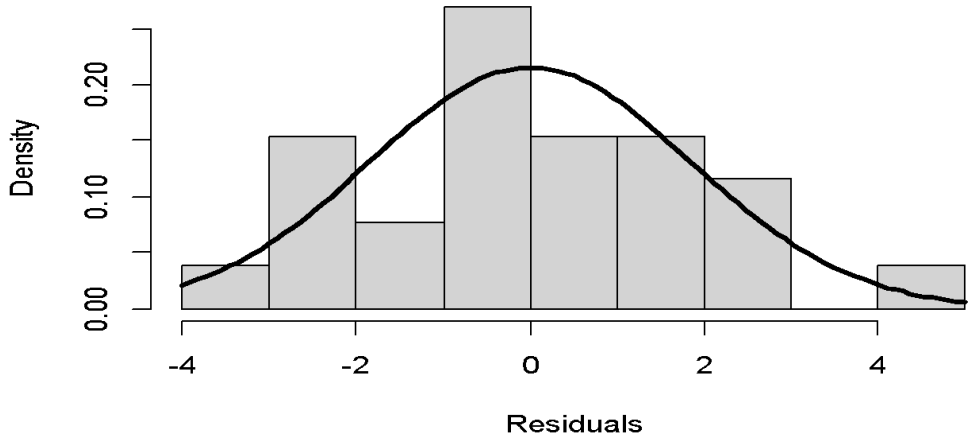

Figure 3: Histogram of Residuals

Forecasting Performance

As *Figure 4* visualizes, forecast plots from the fitted models indicated that the VARX model's independent variables were good predictors of both incidence and mortality rates in the short-term. However, long haul forecasts would require further model refinement as suggested by residual autocorrelation. The mean absolute percentage error (MAPE) for the VARX forecasts was below 20%, indicating acceptable predictive accuracy for public health decision-making despite the noted limitations.

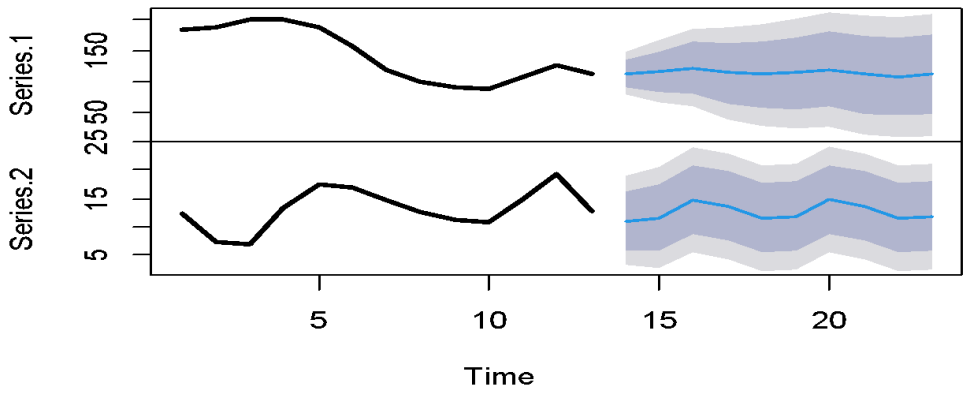

Figure 4: Forecasting Plot

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

DISCUSSION

The analysis sought to capture the interdependencies between incidence and mortality rates through advanced time series techniques in enhancing the understanding of malaria dynamics in Migori County, Kenya. While the dynamics of mortality are influenced, albeit with marginal strength, by both past mortality and exogenous factors, the findings provide evidence that historical values of malaria incidence play a highly significant role in forecasting future incidence. The regression analysis confirmed a significant negative coefficient for ITN use in both the incidence and mortality models, reaffirming the prophylactic benefits of such interventions. However, more variables need to be included in the model owing to the unexpected positive relationship between ITN access and malaria outcomes. The positive relationship could be interpreted as increased access to ITNs reflects a positive public health initiative, lacking correspondence with proper or consistent usage. Further, areas in Migori County with elevated malaria prevalence should be prioritized for the distribution of ITNs and other targeted interventions.

The rich dynamics underlying malaria transmission became clear with the VAR and VARX models in time series analysis. Today, cases of malaria in Migori county are largely determined by recent infection histories, as observed from the significance of lagged terms in the incidence equation, a finding that resonates with the transmission cycle of Plasmodium parasites. VAR and VARX analysis findings match those of Hyndman and Athanasopoulos (2018), where the introduction of exogenous variables in the VARX framework provided more targeted insights into how external interventions and conditions shape the evolution of the epidemic over time.

The BVAR model built on the regression, VAR, and VARX models to incorporate prior distributions, reducing the risk of overfitting in scenarios where the time series data had noise or the sample size was limited. In the present study, the consensus between posterior median estimates and classical OLS estimates lends credibility to our findings, while the Bayesian framework provides a formal mechanism to quantify parameter uncertainty. The use of Monte Carlo simulation via Gibbs sampling cemented the dynamic structure of the data, despite the exclusion of critical environmental, behavioral, and climatic factors (unobserved) that could be included to avert the high residual variance, a major limitation of the study.

Despite the high residual autocorrelation and non-inclusion of environmental, socioeconomic, and behavioral factors, this study provides critical insights for public health policy. The need for community education campaigns should complement the ITNs distribution for their proper and regular use, as the analysis has demonstrated that ITN use significantly mitigates malaria incidence and mortality. Moreover, sustained efforts in early diagnosis and prompt treatment are critical to prevent the escalation of cases, as observed from the strong positive correlation between malaria infection prevalence and incidence. VAR-based models in operational decision-making will enable authorities to anticipate surges and allocate resources more effectively as it offers a dynamic forecasting component of the analysis, despite its need for further refinement.

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

IMPLICATION TO RESEARCH AND PRACTICE

Both academic research and public health practice can derive material implications from this study. The successful application of regression, VAR, VARX, and BVAR models in the modeling of malaria incidence and mortality rates demonstrate the utility of time series techniques in epidemiological data and public health surveillance from a research standpoint. Time series models factor in temporal dependencies and feedback mechanisms inherent in disease transmission, an aspect that many existing studies relying on cross-sectional or univariate models tend to overlook. Our study offers a deeper understanding of disease behavior over time by illustrating how multivariate time series analysis can capture these dynamics. Furthermore, the inclusion of exogenous variables in treatment efficacy and ITN use offers a breakthrough for researchers seeking to construct hybrid models that bridge epidemiological theory with machine learning prediction. Other studies could use the present insights to incorporate higher-dimensional models, spatial-temporal interactions, and climate data for more granular predictive insights.

For public health administrators, practitioners, and policymakers, the present study offers actionable insights worth informing targeted interventions. The importance of promoting the correct and consistent use of nets through behavior change campaigns, besides just distributing them, reaffirms the significant negative relationship between ITN use and both malaria incidence and mortality. Second, the insight on ITN access alone may not significantly mitigate malaria incidence and mortality as accessibility does not amount to utilization, hence disease burden persistence. Therefore, policymakers must think beyond ITNs distribution to include monitoring mechanisms, community education, and stakeholder engagement. The strong predictive power of malaria infection prevalence also emphasizes the need for ongoing surveillance and rapid-response mechanisms to detect and mitigate upswings in transmission. Furthermore, the implementation of real-time VARX-based forecasting tools could allow public health administrators to preempt seasonal surges by dynamically allocating resources such as antimalarial drugs, diagnostic kits, and vector control tools.

Most importantly, our study narrows the methodological rigor of statistical modeling with the urgency and pragmatism of grassroot health initiative, a demonstration that improving incidences of malaria and mortality in low resource settings such as Migori County requires both models and effective implementation.

CONCLUSIONS AND RECOMMENDATIONS

Classical and Bayesian vector autoregressive models are valuable tools in deciphering and forecasting the intricacies of malaria incidence and mortality in Migori County, Kenya, as this study has demonstrated. The study reveals that intervention-related variables (particularly ITN use and effective anti-malaria treatment) play significant roles in adverse outcomes mitigation, while the past malaria incidence is a strong predictor of current cases and future occurrences. However, the complexity of translating public health resources into effective behavioral change comes to play with the finding that ITN access per se may correspond with higher reported cases. As such, we recommend the following for policy development and research:

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

First, there is a need for enhanced data integration. Future policymakers should explore additional exogenous predictors, such as socioeconomic indicators, healthcare infrastructure and access, and climatic factors. Incorporation of additional exogenous variables will fully capture the determinants of malaria dynamics.

Second, future public health interventions should build on VAR, VARX, and BVAR for model refinement. The inclusion of additional lag structures will ensure models address autocorrelation and improve forecast accuracy. Furthermore, refinement of models should consider hybrid models by integrating machine learning techniques or alternative econometric methods, such as the GMM.

Third, the quality of interventions should inform public health policies. There is a need to document and analyze not only the availability of resources but also the quality of implementation of community adherence to recommended practices, given the mixed results observed with ITN access versus ITN use.

While implementing targeted interventions based on our findings, public health administrators should expand geographical scope. These findings could be applicable in counties such as Nyamira with similar demographic and malaria incidence, such as Migori. Comparative analyses across multiple endemic regions would improve the generalizability of the findings and aid in the formulation of tailored regional strategies.

Lastly, we recommend operational forecasting. The dynamic forecasting models developed herein show promise for aiding public health authorities in resource allocation and intervention timing. Efforts to deploy operational versions of these models in real time could add more value in public health management and targeted interventions.

FUTURE RESEARCH

The analysis indicated that VARX performed better than VAR by producing a better fit, despite incidences of autocorrelation in the residuals. The residuals autocorrelation was an indication that model improvement is required in the future through the capture of serial dependencies in the data. Future studies should include environmental and socioeconomic predictors, such as temperature fluctuations, rainfall, level of education, income, health system capacity, etc., as the regression and time series models focus on key interventions. The generalizability of the findings to other counties in Kenya with differing transmission dynamics should be conducted in the future given that the present study relies on administrative data from Migori County, and may not be replicated in other regions.

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

REFERENCES

Achieng, E., Otieno, V., & Mung'atu, J. (2020). Modeling the trend of reported malaria cases in Kisumu County, Kenya. *F1000Research*, *9*, 600. https://doi.org/10.12688/f1000research.23259.1

Badaoui, F., Bouhout, S., Amar, A., & Khomsi, K. (2023). Modelling of Leishmaniasis Infection Dynamics: A Comparative Time Series Analysis with VAR, VECM, Generalized Linear and Markov Switching Models. *Engineering Proceedings*, *39*(1), Article 1. https://doi.org/10.3390/engproc2023039038

Bayarbat, U., & Li, Y. (2020). Empirical Analysis of Relationship between Per Capita Health Expenditure and Economic Growth Based on Vector Autoregressive Model (VAR) in Mongolia. *Theoretical Economics Letters*, *10*(1), Article 1. https://doi.org/10.4236/tel.2020.101010

Chang, L., & Shi, Y. (2024). A discussion on the robust vector autoregressive models: Novel evidence from safe haven assets. *Annals of Operations Research*, *339*(3), 1725–1755. https://doi.org/10.1007/s10479-022-04919-6

Darkoh, E. L., Larbi, J. A., & Lawer, E. A. (2017). A Weather-Based Prediction Model of Malaria Prevalence in Amenfi West District, Ghana. *Malaria Research and Treatment*, *2017*, 7820454. https://doi.org/10.1155/2017/7820454

Diao, O., Absil, P.-A., & Diallo, M. (2023). Generalized Linear Models to Forecast Malaria Incidence in Three Endemic Regions of Senegal. *International Journal of Environmental Research and Public Health*, *20*(13), 6303. https://doi.org/10.3390/ijerph20136303

Eunice, A., Wanjoya, A., & Luboobi, L. (2017). Statistical Modeling of Malaria Incidences in Apac District, Uganda. *Open Journal of Statistics*, *7*(6), Article 6. https://doi.org/10.4236/ojs.2017.76063

Guissou, E., Da, D. F., Hien, D. F. de S., Yameogo, K. B., Yerbanga, S. R., Ouédraogo, G. A., Dabiré, K. R., Lefèvre, T., & Cohuet, A. (2023). Intervention reducing malaria parasite load in vector mosquitoes: No impact on Plasmodium falciparum extrinsic incubation period and the survival of Anopheles gambiae. *PLOS Pathogens*, *19*(5), e1011084. https://doi.org/10.1371/journal.ppat.1011084

Hou, P. S., Fadzil, L. M., Manickam, S., & Al-Shareeda, M. A. (2023). Vector Autoregression Model-Based Forecasting of Reference Evapotranspiration in Malaysia. *Sustainability*, *15*(4), Article 4. https://doi.org/10.3390/su15043675

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). Otexts. https://otexts.com/fpp2/

Karlsson, S., Mazur, S., & Nguyen, H. (2023). Vector autoregression models with skewness and heavy tails. *Journal of Economic Dynamics and Control*, *146*, 104580. https://doi.org/10.1016/j.jedc.2022.104580

Okiro, E. A., Alegana, V. A., Noor, A. M., & Snow, R. W. (2010). Changing malaria intervention coverage, transmission and hospitalization in Kenya. *Malaria Journal*, *9*(1), 285. https://doi.org/10.1186/1475-2875-9-285

Ototo, E. N., Ogutu, J. O., Githeko, A., Said, M. Y., Kamau, L., Namanya, D., Simiyu, S., & Mutimba, S. (2022). Forecasting the Potential Effects of Climate Change on Malaria in the Lake Victoria Basin Using Regionalized Climate Projections. *Acta Parasitologica*, *67*(4), 1535–1563. https://doi.org/10.1007/s11686-022-00588-4

Oyegoke, O. O., Adewumi, T. S., Aderoju, S. A., Tsundzukani, N., Mabunda, E., Adeleke, M. A., Maharaj, R., & Okpeku, M. (2023). Towards malaria elimination: Analysis of travel history and case forecasting using the SARIMA model in Limpopo Province.

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN

Parasitology Research, *122*(8), 1775–1785. https://doi.org/10.1007/s00436-023-07870-y

Savi, M. K. (2023). An Overview of Malaria Transmission Mechanisms, Control, and Modeling. *Medical Sciences*, *11*(1), Article 1. https://doi.org/10.3390/medsci11010003

Shah, I., Mubassir, P., Ali, S., & Albalawi, O. (2024). A functional autoregressive approach for modeling and forecasting short-term air temperature. *Frontiers in Environmental Science*, *12*. https://doi.org/10.3389/fenvs.2024.1411237

Sugita, K. (2022). Forecasting with Bayesian vector autoregressive models: Comparison of direct and iterated multistep methods. *Asian Journal of Economics and Banking*, *6*(2), 142–154. https://doi.org/10.1108/AJEB-04-2022-0044

Sukono, S., Ibrahim, R. A., Riaman, R., Hertini, E., Hidayat, Y., & Saputra, J. (2023). Forecasting the Effectiveness of COVID-19 Vaccination Using Vector Autoregressive with an Exogenous Variable: On the Cases of COVID-19 in Indonesia. *Discrete Dynamics in Nature and Society*, *2023*(1), 6285328. https://doi.org/10.1155/2023/6285328

Warsono, W., Russels, E., Wamiliana, W., Widiarti, W., & Usman, M. (2019). Vector Autoregressive with Exogenous Variable Model and its Application in Modeling and Forecasting Energy Data: Case Study of PTBA and HRUM Energy. *International Journal of Energy Economics and Policy*, *9*(2), Article 2.

Zheng, Y. (2024). An Interpretable and Efficient Infinite-Order Vector Autoregressive Model for High-Dimensional Time Series. *Journal of the American Statistical Association*. https://www.tandfonline.com/doi/abs/10.1080/01621459.2024.2311365

Article DOI: 10.52589/AJMSS_MK6PPMDN
DOI URL: https://doi.org/10.52589/AJMSS_MK6PPMDN