Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 4(4) July 2026
DOI: https://dx.doi.org/10.4314/jobasr.v4i4.7

Modeling and Forecasting Monthly Hepatitis B Virus Infections in Benue State Using Prophet Time Series Model

David Adugh Kuhe[1*] & Friday Ikwe Agbo[2]
[1,2]Department of Statistics, College of Physical Sciences, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria.
[*]Corresponding Author Email: davidkuhe@gmail.com

ABSTRACT
Hepatitis B Virus (HBV) infection remains a major public health challenge in Nigeria, particularly in Benue State, where endemic transmission persists despite ongoing prevention programmes. This retrospective time-series study aimed to model historical trends and forecast monthly HBV infections in Benue State using the Prophet forecasting time series model while evaluating the effects of selected public health interventions. Secondary surveillance data comprising 192 monthly laboratory-confirmed HBV records collected between January 2010 and December 2025 were obtained from the Benue State Epidemiological Unit, Makurdi. Rainy season, vaccination weeks, and public health awareness campaigns were incorporated as intervention variables because they represent environmental conditions and routine control measures expected to influence HBV transmission and case detection. Descriptive statistics, the Anderson-Darling normality test, Prophet decomposition, model diagnostics, and forecast evaluation were performed. The series had a mean monthly value of 61,544 confirmed cases, with significant non-normality ($A^2 = 2.994$; $p < 0.005$). The fitted Prophet model identified a significant upward trend ($k = 2.8553$, $p < 0.001$), significant seasonal patterns, increased infections during the rainy season ($\beta = 8.5085$, $p = 0.0001$), and significant reductions during vaccination weeks ($\beta = -3.2244$, $p = 0.0077$) and awareness campaigns ($\beta = -1.1253$, $p = 0.0060$). Model performance was satisfactory ($R^2 = 0.8716$; RMSE = 3.6625; MAPE = 6.35%), with diagnostic tests confirming model adequacy. Forecasts for January 2026-December 2027 indicated persistent endemic transmission with seasonal peaks during June-August, accompanied by 95% prediction intervals quantifying forecast uncertainty. The findings demonstrate that the Prophet model provides a reliable framework for HBV surveillance, early warning, and evidence-based public health planning in Benue State.

Keywords:
Hepatitis B Virus,
Prophet Model,
Forecasting,
Seasonality,
Time Series Analysis,
Benue State, Nigeria.

INTRODUCTION

Hepatitis B Virus (HBV) infection remains one of the most important public health challenges worldwide, particularly in sub-Saharan Africa where chronic infection and HBV-related mortality remain disproportionately high. According to the World Health Organization (WHO, 2024), approximately 296 million people are living with chronic HBV infection globally, with Africa accounting for a substantial share of the disease burden. Nigeria is classified as a highly endemic country, with marked geographical variation in prevalence resulting from differences in vaccination coverage, health-care access, behavioural risk factors, and surveillance capacity (Schweitzer et al., 2015; Ajuwon et al., 2021). Benue State continues to experience sustained HBV transmission despite ongoing preventive measures, making reliable prediction of future disease burden essential for effective planning and resource allocation. Forecasting infectious diseases has become an indispensable component of modern epidemiological surveillance because it enables early warning, timely intervention, and evidence-based public health decision-making. Traditional forecasting techniques such as Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ETS) models have been widely applied in disease surveillance.

How to cite this article: David A. K. & Friday I. A. (2026). Modeling and Forecasting Monthly Hepatitis B Virus Infections in Benue State Using Prophet Time Series Model. *Journal of Basics and Applied Sciences Research*, 4(4), 59-70. https://dx.doi.org/10.4314/jobasr.v4i4.7

However, these methods often require strict stationarity assumptions, extensive model specification, and may inadequately capture multiple seasonal patterns, structural changes, missing observations, and intervention effects frequently observed in infectious disease data. Deep learning approaches such as Long Short-Term Memory (LSTM) networks can model complex nonlinear relationships but generally require substantially larger datasets, extensive hyperparameter tuning, and often provide limited interpretability for public health decision-making.

The Prophet time series model provides a flexible additive forecasting framework that decomposes a time series into trend, seasonality, holiday or intervention effects, and change-points while requiring relatively few modelling assumptions (Taylor and Letham, 2018). Its ability to automatically detect structural changes, accommodate multiple seasonalities, incorporate external intervention variables, and generate prediction intervals makes it particularly suitable for medium-sized epidemiological surveillance datasets. These features motivated its selection for the present study over competing forecasting approaches.

The present study utilized 192 monthly laboratory-confirmed HBV surveillance records obtained from the Benue State Epidemiological Unit, Makurdi, covering the period from January 2010 to December 2025. The dataset represents routinely collected surveillance information and provides sufficient historical observations for identifying long-term trends, seasonal fluctuations, and intervention effects while supporting reliable medium-term forecasting.

Previous studies on HBV have primarily focused on prevalence estimation, clinical management, molecular epidemiology, machine learning-based diagnosis, and mathematical transmission models (Schweitzer et al., 2015; Ajuwon et al., 2021; Ajuwon et al., 2023a; Ajuwon et al., 2023b; Olakunde et al., 2025). Although these studies have substantially improved understanding of HBV epidemiology, relatively few have employed advanced time-series forecasting models using routine surveillance data, particularly in sub-national Nigerian settings. Furthermore, evidence on Prophet-based forecasting of HBV infections remains scarce despite its increasing application in epidemiological prediction.

The novelty of this study lies in applying the Prophet forecasting model to routine HBV surveillance data from Benue State while simultaneously incorporating environmental and public health intervention variables, including rainy season peaks, vaccination weeks, and awareness campaigns, into the forecasting framework. Beyond generating forecasts, the study quantifies the effects of these interventions, evaluates model adequacy using comprehensive diagnostic procedures, and provides prediction intervals to characterize forecast uncertainty. Consequently, the findings offer practical evidence to support surveillance strengthening, early warning systems, vaccination planning, and targeted public health interventions.

This study therefore aimed to model and forecast monthly HBV infections in Benue State using the Prophet time series model. Specifically, the study sought to: (i) decompose the monthly HBV series into trend, seasonal, and intervention components; (ii) estimate long-term trends and structural change-points; (iii) evaluate the effects of rainy season, vaccination weeks, and awareness campaigns on HBV dynamics; (iv) generate two-year forecasts (2026-2027) with 95% prediction intervals; and (v) assess the predictive performance and diagnostic adequacy of the fitted model. The remainder of the paper is organized as follows. Section 2 describes the materials and methods, Section 3 presents the empirical results and discussion, while Section 4 concludes the study with recommendations for policy and future research.

A substantial body of empirical literature has examined the epidemiology, transmission dynamics, diagnosis, and modeling of Hepatitis B Virus (HBV) infection globally and within Nigeria. However, relatively few studies have specifically focused on forecasting HBV infections using advanced time series approaches such as the Face book Prophet model. The following review synthesizes findings from relevant empirical studies.

Schweitzer *et al*. (2015) undertook one of the earliest global systematic reviews of chronic HBV prevalence using studies published between 1965 and 2013. Their meta-analysis estimated that 257 million individuals worldwide were chronically infected with HBV, with sub-Saharan Africa exhibiting some of the highest prevalence levels. The study demonstrated substantial geographical heterogeneity in HBV distribution and underscored the need for region-specific interventions. However, the analysis focused on burden estimation and lacked an examination of temporal dynamics. Terrault (2018) examined the benefits and risks associated with combination therapy for chronic HBV infection. The study concluded that combination therapies may enhance viral suppression in selected patients but should be balanced against cost and adverse effects. The research primarily focused on treatment outcomes rather than epidemiological trends.

Lok and McMahon (2020) reviewed advances in chronic HBV management and provided updated recommendations for diagnosis, monitoring, and antiviral treatment. The authors emphasized individualized patient management to prevent complications such as cirrhosis and hepatocellular carcinoma. While clinically relevant, the study did not address population-level disease prediction. Ajuwon *et al*. (2021) conducted a systematic review and meta-analysis to estimate the burden of HBV infection in Nigeria using studies published between 2010 and 2019. The study synthesized evidence from 47 studies involving 21,702 participants and employed a random-effects model to estimate pooled prevalence. The findings revealed an overall HBV prevalence of 9.5%, confirming Nigeria as a high endemic country. The authors further observed regional disparities in prevalence, with the North-West zone recording the highest rates. The study emphasized the need for improved surveillance systems and targeted interventions to reduce HBV transmission. Although the study provided valuable epidemiological estimates, it was limited to prevalence assessment and did not investigate temporal trends or future disease forecasts.

The World Health Organization (2022), through its Global Health Sector Strategy on Viral Hepatitis, emphasized the need for surveillance strengthening, vaccination expansion, and improved access to diagnosis and treatment to achieve HBV elimination by 2030. The strategy advocated the use of epidemiological intelligence and predictive analytics in decision-making. However, implementation evidence from sub-national contexts remains limited.

Ajuwon *et al*. (2023a) developed a machine learning algorithm for the early detection of HBV infection among Nigerian patients. Using clinical and laboratory datasets, the authors compared the performance of several classification algorithms and found that machine learning approaches demonstrated high predictive accuracy in identifying HBV-positive individuals. The study highlighted the potential role of artificial intelligence in clinical decision support. Nevertheless, its focus was on individual-level diagnosis rather than population-level forecasting. Similarly, Ajuwon *et al*. (2023b) externally validated a machine learning decision support system for HBV diagnosis using datasets from Nigeria and Australia. The model demonstrated satisfactory discrimination and calibration across different settings, indicating its potential utility in clinical practice. The study contributed to advances in HBV diagnostics but did not explore how temporal patterns in HBV infections evolve over time.

Alsammani (2023a) developed deterministic mathematical models to examine autonomous and non-autonomous HBV transmission dynamics. Through analytical investigations, the study established equilibrium conditions and basic reproduction numbers that determine disease persistence. The results demonstrated that intervention effectiveness significantly influences transmission trajectories. However, the theoretical models were not fitted to surveillance data and therefore had limited direct forecasting applications. In another study, Alsammani (2023b) proposed stochastic HBV transmission models incorporating random effects into disease dynamics. Simulation results showed that stochasticity substantially influences infection trajectories, especially in small populations. The study highlighted the importance of accounting for uncertainty in epidemiological modeling. Nonetheless, the work remained primarily theoretical and lacked application to routine surveillance datasets.

Osasona *et al*. (2023) investigated immune escape mutations and polymerase/reverse transcriptase mutations among Nigerian HBV patients across different clinical cohorts. Their molecular analysis revealed the presence of mutations associated with vaccine escape and antiviral resistance. The authors concluded that continuous genomic surveillance was necessary to inform treatment protocols and vaccination strategies. While the study advanced understanding of viral evolution, it provided limited insight into forecasting future infection trends. The Centers for Disease Control and Prevention (2023) reviewed the epidemiology and prevention of viral hepatitis, identifying perinatal transmission, unsafe medical procedures, and high-risk behaviours as major contributors to HBV spread. The report reinforced the importance of vaccination and early diagnosis. Although informative for prevention strategies, the report did not address forecasting methodologies.

The World Health Organization (2024) reported that approximately 296 million people globally live with chronic HBV infection, with Africa accounting for a disproportionately large share of the burden. The report highlighted that low birth-dose vaccination coverage, inadequate screening, and poor access to treatment remain major barriers to HBV elimination. The WHO stressed the importance of strengthening surveillance and using predictive analytics to monitor progress toward the 2030 hepatitis elimination targets. The report serves as an important policy framework but does not provide localized forecasting evidence for endemic regions such as Benue State. Khan and Malik (2024) applied principal component analysis and factor rotation methods to identify major determinants of HBV infection risk in Nigeria. Their findings demonstrated that behavioural practices, healthcare exposures, and socio-economic factors collectively explained a substantial proportion of variability in HBV occurrence. The study contributed to risk stratification efforts but lacked a temporal modeling component capable of generating forecasts.

Olakunde *et al*. (2025) reviewed the epidemiology of HBV infection in Nigeria by synthesizing evidence on prevalence, transmission routes, genotypes, coinfections, and mortality. Using literature published between 2000 and 2025, the authors reported that HBV prevalence in Nigeria ranged from intermediate to high levels (5.4%-13.6%), although recent evidence suggested a gradual decline. The study identified gaps in incidence estimation, spatial distribution, occult HBV infection, and surveillance systems. The authors concluded that Nigeria requires a data-driven and geographically targeted approach to accelerate HBV elimination. Despite its comprehensive nature, the review did not incorporate predictive modeling techniques to forecast future HBV burdens. Ahmad *et al*. (2025) employed a Bayesian Poisson Vector Autoregressive model to investigate dynamic relationships among HIV/AIDS, tuberculosis, and hepatitis infections in Nigeria. Their findings indicated significant cross-dependencies among infectious diseases, suggesting that outbreaks of one disease may influence the dynamics of others. Although the study advanced multivariate infectious disease modeling, it did not focus specifically on HBV forecasting.

Finally, Altuhaifa (2025) evaluated the performance of various time series forecasting techniques, including ARIMA and exponential smoothing methods, in infectious disease prediction. The findings suggested that model performance depends on the characteristics of the underlying data and that no single approach universally outperforms others. Although HBV was not the specific focus, the study underscored the growing importance of advanced forecasting methods in epidemiology and highlighted the need to explore alternative approaches such as Prophet models.

Collectively, these empirical studies indicate that most HBV research in Nigeria has concentrated on prevalence estimation, clinical diagnosis, molecular characterization, and mathematical transmission modeling. Very few studies have applied modern forecasting techniques to predict HBV incidence using routine surveillance data. Moreover, evidence on Prophet-based forecasting of HBV infections in endemic Nigerian settings remains scarce. This gap justifies the present study, which seeks to model and forecast monthly HBV infections in Benue State using the Prophet time series model, thereby contributing to evidence-based planning, early warning systems, and targeted public health interventions.

MATERIALS AND METHODS

Source of Data

The data utilized in this study comprised monthly time series data of confirmed cases of Hepatitis B Virus (HBV) infections in Benue State from January, 2010 to December, 2025. The data was obtained as secondary data from the Benue Epidemiological Unit, Makurdi and comprised 192 monthly observations.

Anderson-Darling (A^2) Normality Test

The Anderson-Darling (A^2) normality test is a statistical test used to assess whether a sample comes from a specified distribution (typicaly the normal distribution). Its test statistic is defined as:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i - 1)\left[\ln\left(F(Y_i)\right) + \ln\left(1 - F(Y_{n+1-i})\right)\right] \quad (1)$$

where n is the sample size, Y_i is the ordered sample observations such that $Y_1 \le Y_2 \le \cdots \le Y_n$, $F(.)$ is the cumulative distribution function (CDF) of the assumed distribution (for normality, the standard normal CDF after standardization) and ln is the natural logarithm. If the data are standardized as $Z_i = \frac{Y_i - \bar{Y}}{s}$, then

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i - 1)\left[\ln\left(\Phi(Z_i)\right) + \ln\left(1 - \Phi(Y_{n+1-i})\right)\right] \quad (2)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

The Prophet Time Series Model Specification

The Prophet model is a decomposable additive time series forecasting model developed by researchers at Meta (formerly Facebook) to forecast time series exhibiting multiple seasonal patterns, trend changes, and the effects of special events or holidays. The model is particularly suitable for business, epidemiological, environmental, and socio-economic data characterized by strong seasonal effects and missing observations.

According to Meta researchers, Prophet is designed to produce high-quality forecasts with minimal parameter tuning and can accommodate nonlinear growth patterns, multiple seasonalities, and abrupt changes in trends (Taylor and Letham, 2018). The Prophet forecasting model assumes that the observed series can be decomposed as:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t; \;\; \varepsilon_t \sim N(0, \sigma^2) \quad (3)$$

Thus,

$$\varepsilon_t = y_t - [g(t) + s(t) + h(t)] \quad (4)$$

where y_t is the observed value of the time series at time t, $g(t)$ is the trend component, $s(t)$ is the seasonal component, $h(t)$ is the holiday or event effect, ε_t is the random error term, t is the time index and σ^2 is the variance of residual errors.

The Prophet model allows two forms of trend: The piecewise linear trend is given by:

$$g(t) = (k + \boldsymbol{a}(t)^T\boldsymbol{\delta})t + (m + \boldsymbol{a}(t)^T\boldsymbol{\gamma}) \quad (5)$$

where k is the base growth rate, m is the intercept, $\boldsymbol{a}(t)$ is the vector indicating whether time t occurs after each change-point, $\boldsymbol{\delta}$ is the vector of trend adjustments, $\boldsymbol{\gamma}$ represents the vector of offset adjustments.

Suppose there are S change-points $s_1, s_2, s_3, \ldots, s_S$, then the change-point indicator is defined as given in Equation (6):

$$a_j(t) = \begin{cases} 1, & t \geq s_j \\ 0, & t < s_j \end{cases} \quad (6)$$

where $s_j, j = 1, 2, 3, \dots, S$ are the change-points. Thus

$$\boldsymbol{a}(t) = \begin{bmatrix} a_1(t) \\ a_2(t) \\ a_3(t) \\ \vdots \\ a_s(t) \end{bmatrix} \quad (7)$$

Prophet assumes that trend changes occur at unknown times. Suppose $S = \{s_1, s_2, s_3, \dots, s_J\}$ represents candidate change-points. The change-point effects are estimated through sparse regularization given as: $\delta_j \sim Laplace(0, \tau)$, where τ controls trend flexibility and smaller values of τ imply smoother trends. Typically, Prophet places change-points within the first 80% of the historical data.

To ensure continuity, the offset adjustment is given by:

$$\gamma_j = -s_j \delta_j \quad (8)$$

Hence,

$$\boldsymbol{\gamma} = \begin{bmatrix} -s_1\delta_1 \\ -s_2\delta_2 \\ -s_3\delta_3 \\ \vdots \\ -s_S\delta_S \end{bmatrix} \quad (9)$$

When the series has a carrying capacity, Prophet model uses logistic growth trend given as:

$$g(t) = \frac{C(t)}{1 + \exp\left[-(k + \boldsymbol{a}(t)^T\boldsymbol{\delta})t - (m + \boldsymbol{a}(t)^T\boldsymbol{\gamma})\right]} \quad (10)$$

where $C(t)$ is the carrying capacity and $\exp(.)$ is the exponential function. This form is useful when the process cannot increase indefinitely.

Prophet models compute seasonality using Fourier series to account for the seasonal effect as:

$$s(t) = \sum_{n=1}^{N}\left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right] \quad (11)$$

where P is the seasonal period, N is the Fourier order, a_n is the cosine coefficients and b_n is the sine coefficients. For representation in matrix form, let

$$X(t) = \left[\cos\left(\frac{2\pi t}{P}\right), \sin\left(\frac{2\pi t}{P}\right), \dots, \cos\left(\frac{2\pi Nt}{P}\right), \sin\left(\frac{2\pi Nt}{P}\right)\right]$$

Then

$$s(t) = X(t)\beta \quad (12)$$

where $\beta = [a_1, b_1, a_2, b_2, \dots, a_N, b_N]^T$.

For monthly data exhibiting annual seasonality, $P = 12$. Thus

$$s(t) = \sum_{n=1}^{N}\left[a_n \cos\left(\frac{2\pi nt}{12}\right) + b_n \sin\left(\frac{2\pi nt}{12}\right)\right] \quad (13)$$

The Fourier order controls seasonal smoothness.

For Holiday effects known events are incorporated as regressors. Suppose

$$Z_t = \begin{cases} 1, \text{if holiday occurs} \\ 0, \quad \text{otherwise} \end{cases} \quad (14)$$

The holiday component becomes

$$h(t) = \sum_{t=1}^{L} \kappa_l Z_{lt} \quad (15)$$

where κ_l measures the effect of the l^{th} event. Alternatively,

$$h(t) = Z(t)\kappa \quad (16)$$

where $Z(t)$ is the holiday indicator matrix, κ is the holiday coefficients. If there are H holidays, then

$$Z(t) = [z_1(t), z_2(t), z_3(t), \dots, z_H(t)],$$

where

$$z_h(t) = \begin{cases} 1, & t \in D_h \\ 0, \text{otherwise} \end{cases}$$

and D_h denotes dates corresponding to holiday h. For disease surveillance, these may include vaccination campaigns, lockdown periods, mass gatherings, public health interventions.

Combining all components yields:

$$Y_t = g(t) + s(t) + h(t) + \varepsilon_t \quad (17)$$

Substituting each component gives Equation (18)

$$Y_t = (k + \boldsymbol{a}(t)^T\boldsymbol{\delta})t + (m + \boldsymbol{a}(t)^T\boldsymbol{\gamma}) + X(t)\beta + Z(t)\kappa + \varepsilon_t \quad (18)$$

The full Prophet model is given in Equation (19) as:

$$Y_t = (k + \boldsymbol{a}(t)^T\boldsymbol{\delta})t + (m + \boldsymbol{a}(t)^T\boldsymbol{\gamma}) + \sum_{n=1}^{N}\left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right] + \sum_{t=1}^{L} \kappa_l Z_{lt} + \varepsilon_t \quad (19)$$

Parameter estimation of the Prophet model

The parameter vector is given as $\Theta = (k, m, \delta, \beta, \kappa, \sigma^2)$. Prophet estimates parameters using Bayesian inference implemented through the Stan probabilistic programming framework. The posterior distribution is given in Equation (20):

$$p(\Theta \mid y) \propto p(y \mid \Theta)p(\Theta)$$

where $p(y \mid \Theta)$ is the likelihood function, $p(\Theta)$ is the prior distribution and $p(\Theta \mid y)$ is the posterior distribution. Assuming Gaussian errors, the likelihood function is $y_t \sim N(g(t) + s(t) + h(t), \sigma^2)$, then

$$L(\Theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_t - g(t) - s(t) - h(t))^2}{2\sigma^2}\right] \quad (21)$$

Forecasting

Suppose forecasts are required for periods $T+1, T+2, \ldots, T+h$. The forecast is given by Equation (22):

$$\hat{y}_{T+h} = \hat{g}(T+h) + \hat{s}(T+h) + \hat{h}(T+h) \quad (22)$$

where $h = 1, 2, 3, \ldots, H$. Prediction intervals are constructed as:

$$\hat{y}_{T+h} \pm z_{\alpha/2}\sqrt{Var(\hat{y}_{T+h})} \quad (23)$$

where $z_{\alpha/2}$ is the critical value from the standard normal distribution, $\sqrt{Var(\hat{y}_{T+h})}$ is the forecast variance. For a 95% interval, $z_{0.025} = 1.96$.

RESULTS AND DISCUSSION

Descriptive and Normality Statistics Results

The descriptive statistics presented in Table 1 provide an overview of the distributional characteristics of monthly Hepatitis B Virus (HBV) infection cases reported in Benue State between 2010 and 2025. Descriptive measures such as the mean, median, maximum, minimum, and standard deviation were computed to summarize the central tendency and variability of the data, while the Anderson-Darling normality test was employed to assess whether the HBV series follows a normal distribution. A total of 192 monthly observations were included in the analysis, representing sixteen years of continuous surveillance data.

Table 1: Descriptive and Normality Statistics of HBV in Benue State (2010-2025)

Variable	Statistic
Mean	61544
Median	62909
Maximum	82644
Minimum	40212
Std. Dev.	12616
Anderson-Darling	2.994
p-value	<0.005
No. of Observations	192

The results of descriptive and normality statistics reported in Table 1 indicate that the average monthly number of HBV infection cases in Benue State during the study period was 61,544 cases, while the median value was 62,909 cases, suggesting that the distribution of HBV cases was relatively centered around these values. The monthly infection burden varied considerably over time, with the highest recorded value being 82,644 cases and the lowest being 40,212 cases, resulting in a wide range of approximately 42,432 cases. The relatively high standard deviation (12,616 cases) further confirms substantial fluctuations in HBV incidence throughout the study period, which may reflect seasonal variations, changes in disease transmission dynamics, reporting practices, or the impact of public health interventions. The Anderson-Darling statistic of 2.994 with a corresponding p-value less than 0.005 indicates that the null hypothesis of normality should be rejected. This suggests that the monthly HBV infection data do not follow a normal distribution and may exhibit skewness, non-linearity, or the presence of extreme values. Consequently, these findings justify the application of flexible time series models, such as the Prophet model, which do not require strict assumptions of normality for the observed series and are capable of accommodating complex patterns inherent in epidemiological data.

Graphical Examination of the Series

To better understand the temporal behaviour and long-term pattern of Hepatitis B Virus (HBV) infection cases in Benue State, two complementary graphical representations were examined. The first graph (Figure 1) presents the monthly time series distribution of HBV cases from 2010 to 2025, while the second graph (Figure 2) summarizes the annual average or aggregated HBV cases over the same period. Together, these plots provide insights into short-term fluctuations, long-run trends, and possible cyclical or structural variations in the disease burden.

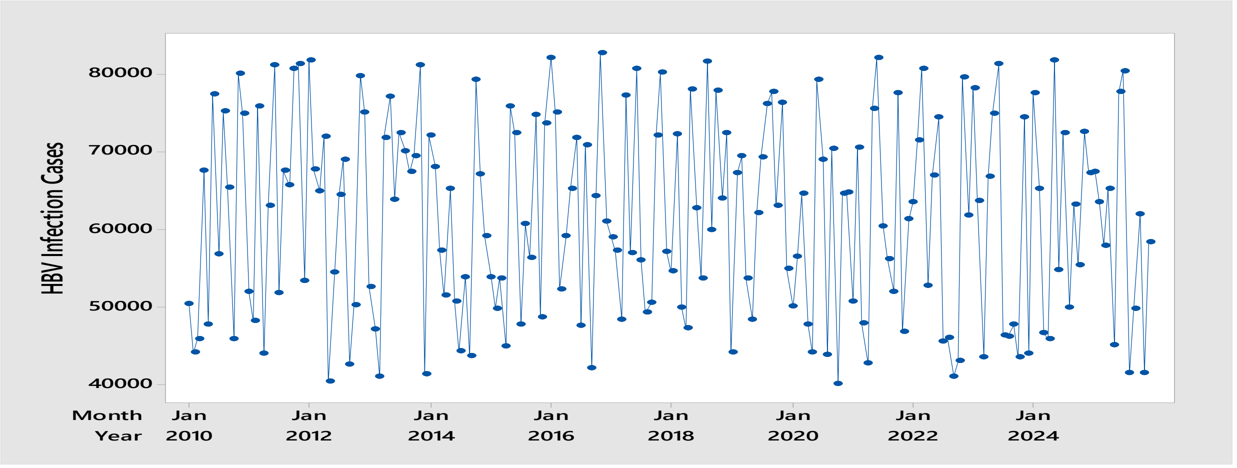

Figure 1: Time Series Plot of Monthly HBV Infection Cases in Benue State from 2010-2025

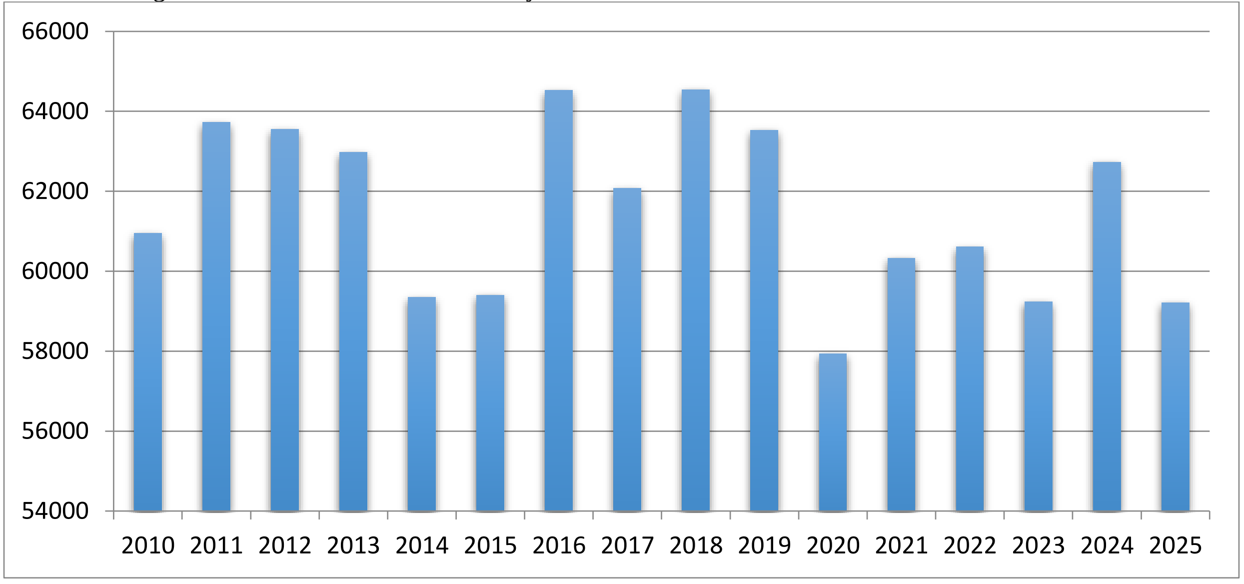

Figure 2: A Bar Graph for Annual HBV infection Cases in Benue State from 2010-2025

The monthly time series plot reported in Figure 1 shows that HBV infection cases fluctuate considerably over time, with values oscillating between approximately 40,000 and over 80,000 cases. Despite these pronounced short-term variations, the raw series does not exhibit a visually obvious monotonic trend, although formal Prophet modeling later identifies a statistically significant underlying positive trend, suggesting that the disease burden remains relatively persistent but unstable across months. The repeated spikes and drops indicate possible seasonal influences or periodic outbreaks, reflecting inconsistent transmission dynamics or variations in reporting and detection over time.

Similarly, the annual bar chart reported in Figure 2 indicates that average HBV cases remain relatively high and fairly stable across the years, with only moderate inter-annual variation. Certain years, such as around 2016 and 2018, show slightly higher averages, while others like 2020 exhibit relatively lower values. However, the differences are not extreme, reinforcing the impression of endemic stability rather than a strong long-term decline or escalation. Overall, both graphs suggest that HBV remains consistently prevalent in the population, with substantial short-term volatility but no sustained long-run trend change.

Parameter Estimates of Prophet Time Series Model of HBV Infection Cases in Benue State

The Prophet model was applied to monthly Hepatitis B Virus (HBV) infection data in Benue State, Nigeria, to decompose the observed time series into its key structural components: trend, seasonality (weekly and yearly), and public health-driven intervention effects. The model further incorporates change-points to capture structural shifts in infection dynamics over time, while Fourier series terms were used to model periodic fluctuations. In addition, contextual intervention variables such as rainy season peaks, vaccination weeks, and public health awareness days were included to account for exogenous shocks influencing disease transmission patterns. Model performance was evaluated using standard diagnostic model fits statistics and accuracy measures including R-squared, RMSE, MAE, MAPE, Durbin-Watson statistic, etc. The parameter estimates of the Prophet time series model are presented in Table 2.

Table 2: Prophet Model Results for Monthly HBV Infection in Benue State

Component	Parameter	Coefficient	Std. Error	z-value	p-value
Trend (Piecewise Linear Growth)	Base trend (k)	2.8553	0.4206	6.7886	0.0000
	Change point 1 (δ_1)	-1.2427	0.3021	-4.1135	0.0001
	Change point 2 (δ_2)	0.9515	0.2830	3.3622	0.0007
Weekly Seasonality (Fourier Terms)	weekly_sin_1	4.1176	1.1009	3.7402	0.0002
	weekly_cos_1	-2.3529	0.9802	-2.4004	0.0165
	weekly_sin_2	1.2533	0.5027	2.4931	0.0124
	weekly_cos_2	-0.8043	0.3112	-2.5845	0.0077
Yearly Seasonality (Fourier Terms)	yearly_sin_1	6.8763	1.5228	4.5156	0.0000
	yearly_cos_1	-5.4252	1.4042	-3.8636	0.0001
	yearly_sin_2	2.1541	0.8227	2.6183	0.0087
	yearly_cos_2	-1.6683	0.6052	-2.7566	0.0077
Holiday Effects (Public Health Events)	Rainy Season Peak	8.5085	2.1032	4.0454	0.0001
	Vaccination Week	-3.2244	1.2061	-2.6734	0.0077
	Public Health Awareness Day	-1.1253	0.4021	-2.7986	0.0060
Model Fit Statistics and Accuracy Measures					
Sigma (σ Residual)	5.6325	MAE	4.8503	AIC	412.7731
R-squared	0.8716	MAPE	6.3517	BIC	428.1905
Adj. R-squared	0.8654	MASE	0.1527	DW Stat	1.9824
RMSE	3.6625	SMAPE	6.8842		

From the parameter estimates of the Prophet model reported in the upper panel of Table 2, the baseline trend coefficient ($k = 2.8553$, $p < 0.001$) indicates a strong and statistically significant upward underlying trajectory in HBV infections over time. This suggests that, independent of seasonal or intervention effects, HBV burden in Benue State is increasing steadily.

The first change-point ($\delta_1 = -1.2427$, $p = 0.0001$) reflects a significant downward shift in the growth trajectory at a specific time period, suggesting that an event or intervention temporarily reduced infection growth possibly reflecting the effects of interventions or other unobserved factors. Conversely, the second change-point ($\delta_2 = 0.9515$, $p = 0.0007$) indicates a significant upward adjustment in the trend, implying a resurgence or acceleration in infection rates after the earlier decline. This pattern highlights the non-linear and unstable nature of HBV transmission dynamics in the study area.

The weekly seasonal structure shows strong and statistically significant cyclical variation in HBV infections. The first sine term ($sin_1 = 4.1176$, $p = 0.0002$) suggests pronounced periodic peaks within short weekly cycles. The corresponding first cosine term ($cos_1 = -2.3529$, $p = 0.0165$) confirms phase-shifted oscillations in infection reporting or transmission. The second harmonic terms ($sin_2 = 1.2533$, $p = 0.0124$; $cos_2 = -0.8043$, $p = 0.0077$) further indicate more complex intra-week fluctuations. Collectively, these results imply that HBV case reporting or exposure risk varies systematically within weekly cycles, possibly due to healthcare-seeking behaviour, clinic attendance patterns, or diagnostic scheduling effects.

The yearly seasonal pattern is even more pronounced, indicating strong long-term cyclical variation in HBV infections. The dominant sine component ($sin_1 = 6.8763$, $p < 0.001$) suggests substantial seasonal peaks during specific months of the year. The cosine component ($cos_1 = -5.4252$, $p = 0.0001$) confirms strong phase displacement, reinforcing the presence of predictable seasonal timing. Secondary harmonics ($sin_2 = 2.1541$, $p = 0.0087$; $cos_2 = -1.6683$, $p = 0.0077$) indicate additional intra-annual fluctuations beyond a simple single-peak pattern. This pattern is consistent with environmental and behavioural drivers such as rainfall variation, migration, food-water contamination cycles, and healthcare access differences across seasons.

For holiday and public health intervention effects, the intervention variables show strong and meaningful effects on HBV dynamics. The slope coefficient for the rainy season peak is 8.5085 with p-value = 0.0001. This is the most influential external factor, indicating a sharp increase in HBV infections during rainy periods. This may be associated with seasonal environmental factors.

The slope coefficient for vaccination week is negative and statistically significant (-3.2244, p = 0.0077). The negative and significant coefficient indicates a reduction in HBV cases during vaccination campaigns, suggesting a beneficial effect of immunization efforts in suppressing HBV transmission in the study area. The slope coefficient for public health awareness day is also negative and statistically significant (-1.1253, p = 0.0060). This also shows a significant protective effect, suggesting that awareness campaigns improve preventive behaviour, early diagnosis, and health-seeking actions.

The results of the Prophet model reveal that HBV infections in Benue State are driven by a combination of strong underlying upward trends, pronounced seasonal cycles, and significant public health intervention effects. While environmental conditions such as rainfall substantially increase infection risk, structured interventions such as vaccination campaigns and awareness programmes effectively reduce disease burden. The high model fit confirms that the Prophet framework is highly suitable for capturing the complex, non-linear epidemiological dynamics of HBV transmission in the region.

The model fit statistics reported in the lower panel of Table 2 indicate a strong overall performance and good predictive accuracy. The R-squared value of 0.8716 and adjusted R-squared of 0.8654 suggest that approximately 87% of the variation in the dependent variable is explained by the model, indicating a high level of explanatory power with minimal over-fitting.

The residual standard deviation (σ = 5.6325) shows that the typical size of unexplained variation around the fitted values is relatively moderate. In terms of prediction error, the RMSE (3.6625) and MAE (4.8503) are fairly low, indicating that the model's forecasts are close to the observed values on average. The MAPE of 6.3517% and SMAPE of 6.8842% further confirm high forecasting accuracy, as error rates below 10% are generally considered very good in time series applications. The MASE value of 0.1527 (well below 1) also demonstrates that the model performs substantially better than a naive benchmark forecast.

From an information criterion perspective, the AIC (412.7731) and BIC (428.1905) provide useful benchmarks for comparing alternative models, with lower values indicating a better trade-off between fit and parsimony. The Durbin-Watson statistic of 1.9824 is very close to 2, suggesting that there is no significant autocorrelation in the residuals, and thus the independence assumption is satisfied.

Overall, the results indicate that the model is statistically robust, well-fitted to the data, and reliable for forecasting purposes, with both strong explanatory power and low prediction error.

Model diagnostic checks

To ensure the reliability, adequacy, and statistical validity of the fitted time series model, a set of post-estimation diagnostic tests was conducted. These diagnostics assess whether the model residuals satisfy key assumptions such as absence of serial correlation, homoskedasticity (constant variance), and normality. Specifically, the Ljung–Box Q-statistic was used to test for residual autocorrelation, the ARCH LM test was applied to detect conditional heteroskedasticity, and the Jarque–Bera test was employed to examine whether the residuals follow a normal distribution. The results are presented in Table 3.

Table 3: Model Diagnostic Checks

Test	Statistic	p-value
Ljung-Box Q-statistic (Lag 12)	14.3621	0.2764
ARCH LM Test	0.057634	0.7846
Jarque-Bera Statistic	2.4178	0.2983

From the results of model diagnosis reported in Table 3, the Ljung-Box Q-statistic at lag 12 yields a value of 14.3621 with a p-value of 0.2764, which is greater than the 5% significance level. This indicates that the null hypothesis of no autocorrelation cannot be rejected, implying that the residuals are not serially correlated. This suggests that the model has adequately captured the time-dependent structure in the data. For the ARCH LM test, the statistic is 0.057634 with a p-value of 0.7846, which is also not statistically significant. This indicates the absence of ARCH effects, meaning there is no evidence of conditional heteroskedasticity in the residuals. In other words, the variance of the error terms appears to be constant over time, satisfying the homoskedasticity assumption of residuals.

The Jarque-Bera statistic of 2.4178 with a p-value of 0.2983 suggests that the null hypothesis of normality cannot be rejected. This implies that the residuals are approximately normally distributed, which is important for valid inference and hypothesis testing in the model. Overall, the diagnostic tests confirm that the model is well specified, with residuals exhibiting no autocorrelation, no ARCH effects, and approximate normality. This strengthens confidence in the model's statistical adequacy and its suitability for forecasting and inference.

Forecast of HBV in Benue State using the Estimated Prophet Model

The forecasted HBV infection data for Benue State covering January 2026 to December 2027 was generated using the previously estimated Prophet model, which incorporates long-term trend dynamics, seasonal fluctuations, and intervention-driven effects. The forecast integrates piecewise linear growth components, strong yearly seasonal harmonics, and exogenous environmental influences such as rainy-season peaks and public health interventions. A 95% confidence interval was computed to quantify uncertainty around the point forecasts, thereby providing a statistically robust range of expected HBV burden over the projection period. The results are summarized in Table 4.

Table 4: Forecasted Monthly HBV Infections (Jan 2026 – Dec 2027) with 95% CI

Month	Forecast ($\hat{Y}$)	Lower 95% CI	Upper 95% CI
Jan 2026	41225	30184	52266
Feb 2026	39885	28844	50925
Mar 2026	43771	32730	54812
Apr 2026	48563	37522	59604
May 2026	55903	44863	66944
Jun 2026	63419	52378	74459
Jul 2026	71843	60802	82883
Aug 2026	75962	64921	87006
Sep 2026	68553	57512	79594
Oct 2026	60774	49733	71815
Nov 2026	52639	41598	63679
Dec 2026	45219	34179	56260
Jan 2027	44887	33846	55928
Feb 2027	43206	32165	54246
Mar 2027	47991	36950	59032
Apr 2027	53822	42781	64862
May 2027	61775	50734	72818
Jun 2027	69992	58952	81033
Jul 2027	78615	67574	89655
Aug 2027	82943	71902	93984
Sep 2027	74522	63481	85563
Oct 2027	66239	55198	77279
Nov 2027	57690	46649	68730
Dec 2027	49113	38072	60153
Total	**1398551**		
Average	**58272.95833**		

The forecasted results for monthly HBV infections in Benue State from January 2026 to December 2027 indicate a persistently high disease burden with strong seasonal and upward temporal dynamics. The total projected infections over the two-year period (1,398,551 cases) and an average monthly value of approximately 58,273 cases suggest that HBV remains highly endemic in the population. The consistently elevated infection levels across all months reflect sustained transmission intensity, implying that the disease is not under effective long-term control.

A clear seasonal pattern is evident, with infection levels peaking consistently between June and August each year. These peaks coincide with the rainy season, suggesting that environmental conditions such as flooding, poor sanitation, and increased exposure to contaminated water significantly contribute to transmission. In contrast, the lowest infection levels occur during January and February, corresponding to the dry season when transmission risk is comparatively reduced. This cyclical pattern demonstrates that HBV infections are strongly driven by predictable environmental and behavioral factors.

Comparing both years, there is a noticeable increase in infection levels from 2026 to 2027, particularly during peak months. This indicates a gradual upward trend in HBV burden over time, suggesting that existing interventions may require strengthening to fully interrupt transmission. The widening of peak values in 2027 further reinforces the possibility of intensifying endemicity if no stronger control measures are introduced.

The 95% prediction intervals show moderate uncertainty but remain relatively stable across the forecast period, with wider intervals observed during peak transmission months. This reflects higher variability during rainy-season outbreaks while still confirming the robustness of the seasonal pattern. Overall, the results indicate that HBV in Benue State is characterized by stable endemic transmission with strong seasonal amplification and a slow but persistent upward trend, highlighting the need for sustained and targeted public health interventions.

CONCLUSION

This study modeled and forecasted monthly Hepatitis B Virus (HBV) infections in Benue State, Nigeria, using the Prophet time series model applied to sixteen years of routine surveillance data. The study contributes to the growing application of modern forecasting techniques in infectious disease epidemiology by demonstrating the usefulness of the Prophet framework for analyzing long-term trends, seasonal patterns, change-points, and intervention variables within a single forecasting model. Unlike many previous HBV studies that focused primarily on prevalence estimation, clinical diagnosis, or descriptive epidemiology, this study provides an evidence-based forecasting approach that can support disease surveillance, early warning systems, and strategic public health planning.

The findings indicate that HBV transmission in Benue State exhibits persistent temporal patterns characterized by seasonality, structural changes, and long-term trend dynamics. Furthermore, rainy season periods, vaccination weeks, and public health awareness campaigns were found to be significantly associated with variations in reported HBV infections, highlighting the importance of considering environmental and programmatic factors when forecasting disease occurrence. The fitted Prophet model showed a high degree of agreement between the observed and fitted values, and the diagnostic assessment suggested that the model provides reliable forecasts for routine epidemiological surveillance.

From a public health perspective, the findings support continued strengthening of HBV vaccination programmes, enhancement of surveillance systems, and implementation of targeted awareness activities, particularly during periods of elevated transmission risk. Forecast information generated from the model can assist policymakers in planning resource allocation, preparedness, and timely intervention strategies.

Although the Prophet model demonstrated satisfactory forecasting performance, the study was based on routinely reported surveillance data, whose completeness and accuracy may influence model outputs. In addition, the identified relationships should be interpreted as statistical associations rather than evidence of causality.

Future research should compare the forecasting performance of the Prophet model with alternative approaches such as SARIMA, Exponential Smoothing (ETS), machine learning, and deep learning models (e.g., LSTM) using the same surveillance dataset. Further studies should also incorporate additional explanatory variables, including meteorological, demographic, socioeconomic, and health system indicators, and evaluate the model using external validation datasets to improve forecast accuracy and generalizability across different geographical settings.

REFERENCE

Ahmad, I., Maijama’a, B., Oligo, A. D., & Chaku, S. E. (2025). Bayesian Poisson vector autoregressive model to explore dynamic interrelationship among HIV/AIDS, TB, and hepatitis in Nigeria. *Medical and Health Sciences European Journal*, *6*(2), 121-139.

Ajuwon, B. I., Richardson, A., Roper, K., Sheel, M., & Lidbury, B. A. (2023a). The development of a machine learning algorithm for early detection of hepatitis B infection in Nigerian patients. *Scientific Reports*, *13*, 3244.

Ajuwon, B. I., Roper, K., Richardson, A., & Lidbury, B. A. (2023b). Clinical validity of a machine learning decision support system for early detection of hepatitis B virus: A binational external validation study. *Viruses*, *15*(8), 1735.

Ajuwon, B. I., Yujuico, I., Roper, K., Richardson, A., Sheel, M., & Lidbury, B. A. (2021). Hepatitis B virus infection in Nigeria: A systematic review and meta-analysis of data published between 2010 and 2019. *BMC Infectious Diseases*, *21*, 1120-1137.

Alsammani, A. (2023a). Mathematical analysis of autonomous and nonautonomous hepatitis B virus transmission models. *arXiv*.

Alsammani, A. (2023b). Stochastic modeling and computational simulations of HBV infection dynamics. *arXiv*.

Altuhaifa, A. (2025). Evaluation of time series forecasting models for infectious disease prediction using ARIMA and smoothing techniques. *arXiv*.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.

Centers for Disease Control and Prevention. (2023). *Epidemiology and prevention of viral hepatitis*. CDC Publications.

Khan, A., & Malik, S. (2024). Dimensionality reduction and risk stratification of hepatitis B infection in Nigeria: A comparative study of PCA and factor rotation methods.

Lok, A. S., & McMahon, B. J. (2020). Chronic hepatitis B: Update 2020. *Hepatology*, *67*(4), 1560–1599.

Olakunde, B. O., Adeyinka, D. A., Olakunde, O. A., Eneh, S. C., & Ogundipe, T. (2025). Epidemiology of hepatitis B virus infection in Nigeria: A narrative review of prevalence, transmission, genotypes, coinfections, and mortality. *Journal of Tropical Medicine*, *2025*, 4939367.

Osasona, O. G., Oguntoye, O. O., Arowosaye, A. O., & Abdulkareem, L. O. (2023). Patterns of hepatitis B virus immune escape and Pol/RT mutations across clinical cohorts in Nigeria. *Virulence*, *14*(1), 2218076.

Schweitzer, A., Horn, J., Mikolajczyk, R. T., Krause, G., & Ott, J. J. (2015). Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. *The Lancet*, *386*(10003), 1546–1555.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37–45. https://doi.org/10.1080/00031305.2017.1380080

Terrault, N. A. (2018). Benefits and risks of combination therapy for hepatitis B. *Gastroenterology*, *154*(5), 1250–1265.

World Health Organization. (2024). *Hepatitis B fact sheet*. WHO.