DEPARTMENT OF STATISTICS
AHMADU BELLO UNIVERSITY, ZARIA
2nd International Conference on Statistical Sciences and Data Analytics (ICSSDA 2026) Proceedings Paper

A Catalytic Linear Infection Model for Age-Structured Hepatitis Data: Estimation and Application to Plateau State, Nigeria

Kenret Danjan[2*], Jamila Abdullahi[1], Ibrahim Abubakar Sadiq[1]

[1]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria
[2]Plateau State Bureau of Statistics, Jos, Nigeria.
*Correspondence: kenretdanjan@gmail.com

Abstract

Hepatitis remains a major public health concern in Nigeria, yet quantitative evidence on its age-specific transmission dynamics is limited. This study applies the Catalytic Linear Infection Model (CLIM) to characterise the age-dependent force of infection of hepatitis in Plateau State, Nigeria. Age-stratified hepatitis surveillance data were obtained from the Surveillance, Outbreak Response, Management and Analysis System (SORMAS) of the Plateau State Ministry of Health, while population data were sourced from the National Population Commission through the Plateau State Bureau of Statistics. Model parameters were estimated using the maximum likelihood method, and their precision was assessed through bootstrap procedures. The estimated force of infection increased linearly after a threshold age of approximately 1.5 years, with a mean time to infection of 12.13 years (95% CI: 12.12–12.15 years). These findings establish the CLIM as a robust framework for modelling age-specific hepatitis transmission and provide valuable epidemiological evidence to support age-targeted prevention and control strategies. The study also extends the application of catalytic modelling to hepatitis epidemiology in Nigeria.

Keywords: Catalytic Linear Infection Model; Hepatitis Transmission; Age-structured Data; Maximum Likelihood Estimation; Mean Time to Infection; Simulation Study; Plateau State, Nigeria.

1. Introduction

The force of infection (FOI) is one of the most important parameters in infectious disease modelling because it quantifies the rate at which susceptible individuals become infected (Saikh *et al*., 2021). Since infection times are rarely observed directly, statistical models are required to infer transmission dynamics from age-specific disease data. Catalytic models have become popular for this purpose because they relate observed infection patterns to the underlying hazard of infection. (Falodun *et al*., 2025)

Infectious diseases continue to impose substantial public health burdens globally, particularly in low- and middle-income countries where socioeconomic conditions, limited access to healthcare, and infrastructural challenges facilitate disease transmission (Falodun *et al*., 2025). Among these, hepatitis infections encompassing hepatitis A, B, and C remain critical causes of morbidity and

mortality in Nigeria. The chronic sequelae of viral hepatitis, including cirrhosis and hepatocellular carcinoma, make it an urgent target for surveillance and intervention (Olakunde *et al*., 2025).

Understanding how hepatitis infections spread through populations and across age groups is fundamental to designing effective prevention and control strategies. Epidemiological modelling offers powerful tools for quantifying disease transmission dynamics, especially when direct longitudinal data are unavailable (Kwon & Lee, 2011). In this context, catalytic infection models provide a mathematical framework to estimate the force of infection (FOI), the rate at which susceptible individuals acquire infection, based on age-specific prevalence data (Kamau et al., 2025; Grosso et al., 2025; Blaizot et al., 2019).

The classical catalytic model assumes a constant infection hazard across all ages. However, exposure to infectious diseases often changes with age due to behavioural, environmental, and biological factors. (Lombard *et al*., 2025) The Catalytic Linear Infection Model (CLIM), proposed by Griffiths (1974), relaxes this assumption by allowing the force of infection to increase linearly after a threshold age.

Although hepatitis remains a major public health challenge in Nigeria, most previous studies have focused on prevalence estimation or risk-factor analysis (Olakunde *et al*., 2025). Limited attention has been given to estimating the age-dependent force of infection using likelihood-based catalytic models. This study therefore applies the CLIM to age-structured hepatitis data from Plateau State to estimate transmission parameters and evaluate its statistical performance.

Most existing hepatitis studies estimate prevalence and associated risk factors but provide limited information on how infection risk changes with age. Traditional catalytic models also assume constant infection pressure, which may not reflect the true transmission process. (Wariri *et al*., 2025) Consequently, there is a need for a flexible statistical model capable of estimating age-dependent force of infection using routinely collected surveillance data.

To estimate the age-dependent force of infection of hepatitis using the Catalytic Linear Infection Model and evaluate its performance using age-structured surveillance data from Plateau State, Nigeria. The objectives of this study are to estimate model parameters using Maximum Likelihood Estimation, to estimate the Mean Time to Infection and to demonstrate the applicability of the model using Plateau State hepatitis data.

The study contributes to statistical modelling by providing a likelihood-based framework for estimating age-dependent force of infection. The model also provides useful evidence for designing age-targeted hepatitis control strategies.

2. Materials and Methods.

2.1. Study Area and Data

Age-structured hepatitis surveillance data for Plateau State, Nigeria, were analysed. The data consisted of reported hepatitis cases grouped by age categories and obtained from routine surveillance records for 2022. The model estimates transmission parameters from age-specific infection frequencies.

Griffiths (1974) defined the CDF and PDF of a catalytic linear infection model as:

$$F(t) = 1 - \exp\left\{\frac{1}{2}a\left((\tau + c)^2 - (t + c)^2\right)\right\}, \; t > \tau \tag{1}$$

$$f(t) = a(t + c)\exp\left\{\frac{1}{2}a\left[(\tau + c)^2 - (t + c)^2\right]\right\}, \quad t > \tau \tag{2}$$

where $a > 0$ represents the slope of the hazard, $c > 0$ adjusts the intercept, and $\tau > 0$ is the threshold age below which infection risk is negligible.

Griffiths (1974) also noted that for equations (1) and (2), the force of infection for children's diseases at least aged between zero and adolescent age is approximated by a linear function given as:

$$\psi(t)=\begin{cases}a(t+c), & t>\tau\\ 0, & t\le\tau\end{cases} \tag{3}$$

where $\psi(t)$ is the force of infection at time t, conditional on not having been infected before. The CLIM model in equation (3) assumes that below a certain time (say, before a child reaches some age τ), there is no risk of infection. This reflects immunity (maternal antibodies, vaccination, or biological immaturity). So we set $\psi(t)=0$ for $t\le\tau$. After τ, the hazard increases linearly with age. That is, the risk of infection per unit time grows proportionally with t, with slope a. The shift parameter c ensures positivity and allows flexibility. Therefore, the CLIM model, $\psi(t)=a(t+c)$ for $t>\tau$ in equation (3), is a simple linear-in-age force of infection.

2.2. Survival Function

The survival function, S(t), gives the probability of not being infected by time t (Sadiq *et al*., 2022; Sadiq *et al*., 2023a; Sadiq *et al*., 2023b; Sadiq *et al*., 2023c). From equation (1),

$$S(t)=1-F\left(t\right)=\exp\left\{\frac{1}{2}a\left[(\tau+c)^2-(t+c)^2\right]\right\},\quad t>\tau \tag{4}$$

This function decays exponentially, meaning the probability of avoiding infection decreases over time.

2.3. Hazard Function

The hazard function h(t) describes the instantaneous rate of infection at time t, given survival up to that point (Sadiq *et al*., 2025; Sadiq *et al*., 2024; Sadiq *et al*., 2025a):
From equations (2) and (4)

$$h(t)=\frac{f(t)}{S(t)}=\frac{a(t+c)\exp\left\{\frac{1}{2}a\left[(\tau+c)^2-(t+c)^2\right]\right\}}{\exp\left\{\frac{1}{2}a\left[(\tau+c)^2-(t+c)^2\right]\right\}}=a(t+c),\quad t>\tau \tag{5}$$

The results in equation (5) show that the infection risk grows linearly over time.

2.4. Statistical Estimation of CLIM Parameters

2.4.1. Likelihood-based estimation

Assume independent observations $(t_1,\ldots,t_n)$ with all $t_i>\tau$ andcontinuous-time (individual times). The likelihood is the product of the PDF $f(t)$ values for each observed time (Sadiq *et al*., 2022; Sadiq *et al*., 2023a; Sadiq *et al*., 2023b; Sadiq *et al*., 2023c). The likelihood function of CLIM is derived by using equation (32) as follows:

$$L(a,c)=\prod_{i=1}^{n}f(t_i)=\prod_{i=1}^{n}\left[a(t_i+c)\exp\left\{\frac{a}{2}((\tau+c)^2-(t_i+c)^2)\right\}\right] \tag{6}$$

The Log-likelihood function of the CLIM is obtained from equation (33) as follows:

$$\ell(a,c)=n\log a+\sum_{i=1}^{n}\log(t_i+c)+\frac{a}{2}n(\tau+c)^2-\frac{a}{2}\sum_{i=1}^{n}(t_i+c)^2 \tag{7}$$

Taking the partial derivative of equation (3) with respect to the parameter (a), we obtain the following:

$$\frac{\partial \ell}{\partial a} = \frac{n}{a} + \frac{1}{2} n(\tau + c)^2 - \frac{1}{2}\sum_{i=1}^{n}(t_i + c)^2 \tag{8}$$

Setting equation (35) to zero and solving for the parameter (a), we have:

$$\frac{n}{a} = \frac{1}{2}\left(\sum_{i=1}^{n}(t_i + c)^2 - n(\tau + c)^2\right) \tag{9}$$

Hence, a closed-form estimator for the parameter $(\hat{a})$ from equation (36) is given as:

$$\hat{a} = \frac{2n}{\sum_{i=1}^{n}(t_i + c)^2 - n(\tau + c)^2} \text{ ; (provided denominator (>0)).} \tag{10}$$

Taking the partial derivative of equation (34) with respect to the parameter (c), we obtain the following:

$$\frac{\partial \ell}{\partial c} = \sum_{i=1}^{n}\frac{1}{t_i + c} + an(\tau + c) - a\sum_{i=1}^{n}(t_i + c) \tag{11}$$

2.4.2. Bootstrap and Delta Methods

To estimate the uncertainty (standard errors and confidence intervals) around the MTI and the parameters (a, c) in CLIM, researchers cannot simply use closed-form formulas because the MTI is a complex non-linear function of a and c.

In CLIM applications (like the Hepatitis seroprevalence study mentioned earlier), the Bootstrap and Delta method are the two standard frequentist approaches used to quantify this uncertainty.

3. Results and Discussions

The results obtained from the application of the CLIM to hepatitis transmission data collected from Plateau State, Nigeria, for the year 2022. The analysis focuses on estimating key epidemiological parameters, including the force of infection, threshold age, and the MTI, using MLE techniques.

2.4. Application to Hepatitis Data

Table 1 provides the parameter estimates for the hepatitis transmission model in Plateau State, including bootstrap standard errors and the mean time to infection with its corresponding 95% confidence interval.

Table1. Parameter estimates for the CLIM

Parameter	Estimates	Bootstrap Standard Errors
$\hat{a}$	0.009636604	7.786835e-06
$\hat{c}$	0.963660385	1.956268e-09
$\hat{\tau}$	1.499339432	8.666742e-10
Mean Time to Infection	12.13374	DM 95% CI-MTI [12.1198, 12.14767]

Table 1 presents the maximum likelihood estimates and bootstrap standard errors for the parameters (a,c, and τ) of the CLIM fitted to the 2022 hepatitis data from Plateau State. The estimated slope parameter $\hat{a} = 0.00964$ indicates that the force of infection increases gradually with age, suggesting a slow but steady rise in hepatitis exposure risk as individuals grow older. The estimate

of the intercept-shifting parameter $\hat{c} = 0.96366$ is small and precisely estimated (bootstrap SE on the order of 1.956268e-09, implying minimal deviation from a near-linear age-infection relationship when ages are shifted by c. The thresholdparameter $\hat{\tau} = 1.4993$ years indicates that the model identifies a short "latency window" in early childhood before the force of infection begins to increase measurably. This is consistent with public health patterns where hepatitis infections in Plateau State are relatively rare among infants but begin to accumulate more noticeably from early childhood onward.

The Mean Time to Infection (MTI), estimated at 12.13 years with a narrow delta-method 95 per cent confidence interval [12.12, 12.15], provides a population-level summary of the average waiting time until infection occurs under the CLIM assumptions. The precision of this estimate, driven by the very small standard errors of the model parameters, reflects a stable age-infection structure within the surveillance data. Biologically, an MTI of approximately 12 years suggests that hepatitis exposure in Plateau State accumulates slowly, with the majority of infections occurring in late childhood and adolescence rather than early in life. This pattern aligns with the observed epidemiological context in which hepatitis transmission is primarily driven by behavioural and environmental risk factors that become more relevant as individuals age. The tight confidence interval further indicates that the CLIM provides internally consistent estimates of transmission timing.

A useful perspective on the Plateau hepatitis estimates is obtained by comparing them with the classic catalytic analysis of measles transmission presented in Griffiths (1974). In Griffiths' work, the estimated force of infection parameters indicated a rapidly increasing infection hazard during early childhood, with measles infections typically occurring within the first few years of life. Specifically, Griffiths' fitted linear catalytic models produced high values of the slope parameter (a) and very small threshold ages (ĉ), reflecting the strong epidemic pressure of measles and its early-age transmission pattern. Consequently, the mean age at infection for measles, often between 3 and 5 years in pre-vaccination settings, was much lower than the MTI observed for hepatitis in Plateau State. The steep rise in measles FOI in Griffiths' model contrasts sharply with the slow, gradual increase captured in the Plateau hepatitis data.

In contrast, the CLIM fitted to the hepatitis data yields a much smaller slope parameter and a mean time to infection of approximately 12.13 years, indicating a far more protracted accumulation of risk across the lifespan. Whereas Griffiths' measles model captures a vivid early-age infection spike, the Plateau hepatitis model reflects a disease with lower transmissibility, more heterogeneous exposure pathways, and greater dependence on behavioural, environmental, or iatrogenic factors that become relevant in later childhood and adolescence. Moreover, the estimated threshold parameter years suggests a modest delay before meaningful exposure occurs, again consistent with hepatitis but very different from measles, where infection pressure is present almost from birth in highly endemic populations. These differences highlight the flexibility of catalytic models but also emphasise that infection-age structures are deeply disease-specific. CLIM, when applied to hepatitis, produces a smooth and gradual FOI curve, whereas the same model applied to a classic childhood infection, such as measles, as in Griffiths (1974), captures a sharply accelerated early-life infection process. This comparison gives the appropriateness of catalytic modelling for a broad class of infections while illustrating how parameter contrasts reveal fundamental differences in disease ecology and transmission dynamics.

4. Conclusion

The findings demonstrate that the CLIM offers a superior and epidemiologically meaningful framework for understanding hepatitis transmission dynamics in Plateau State. The model's ability to incorporate a threshold age and a linearly increasing force of infection aligns well with the observed data, capturing both early-childhood susceptibility and the gradual accumulation of infection risk through adolescence**.**

The estimated MTI of approximately 12 years provides important insight into the typical age at which individuals acquire hepatitis infection in this setting.

Importantly, the strong performance of CLIM in real-world data indicates that it is a reliable and interpretable tool for age-structured infectious disease modelling. The study extends the foundational work of Griffiths by demonstrating that catalytic modelling remains relevant and effective for contemporary epidemiological challenges, especially where disease exposure increases progressively with age. The results call attention to the need for timely, age-targeted interventions and support the integration of catalytic models into hepatitis surveillance and planning in Nigeria.

References

Blaizot, S., Herzog, S. A., Abrams, S., Theeten, H., Litzroth, A., & Hens, N. (2019). Sample size calculation for estimating key epidemiological parameters using serological data and mathematical modelling. *BMC Medical Research Methodology, 19*(1), 51.

Falodun, M. O., Olorunfemi, O., & Irinoye, O. O. (2025). Infectious diseases: Addressing global challenges and prevention strategies for national health improvement. *Community-Acquired Infection, 12*. https://doi.org/10.54844/cai.2024.0533

Gavi, the Vaccine Alliance. (2019). *Childhood immunisation in low- and middle-income countries: Progress and challenges*. https://www.gavi.org

Griffiths, D. A. (1974). A catalytic model of infection for measles. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 23*(3), 330–339.

Grosso, A., Hens, N., & Abrams, S. (2025). An integrative review of the combined use of mathematical and statistical models for estimating malaria transmission parameters. *Malaria Journal, 24*(1), 173.

Hagan, O. C., Nsiah, P., Obiri-Yeboah, D., Yirdong, F., Annan, I., Eliason, S., et al. (2018). Impact of universal childhood vaccination against hepatitis B in Ghana: A pilot study. *Journal of Public Health in Africa, 9*(2), 721.

Kamau, E., Chen, J., Bajaj, S., Torres, N., Creswell, R., Pavlich-Mariscal, J. A., … Lambert, B. (2025). The mathematics of serocatalytic models with applications to public health data. *Statistics in Medicine, 44*(15–17), e70188.

Kwon, S. Y., & Lee, C. H. (2011). Epidemiology and prevention of hepatitis B virus infection. *The Korean Journal of Hepatology, 17*(2), 87–95.

Lombard, B., Cohen, C., von Gottberg, A., Dushoff, J., Pulliam, J. R., & van Schalkwyk, C. (2025). A generalised catalytic model to assess changes in risk for multiple reinfections with SARS-CoV-2. *PLOS ONE, 20*(1), e0315476.

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*(1), 4939367.

Sadiq, I. A., Doguwa, S. I. S., Yahaya, A., & Garba, J. (2023c). New Generalised Odd Fréchet-Odd Exponential-G Family of Distributions with Statistical Properties and Applications. FUDMA Journal of Sciences, 7(6), 41-51. [https://doi.org/10.33003/fjs- 2023-0706-2096]

Sadiq, I. A., Doguwa, S. I. S., Yahaya, A., & Usman, A. (2023b). Development of New Generalised Odd Fréchet-Exponentiated-G Family of Distribution. UMYU Scientifica, 2(4), 169-178. [https://doi.org/10.56919/usci.2324.021]

Sadiq, I. A., Doguwa, S. I., Yahaya, A., & Garba, J. (2022). New Odd Fréchet-G Family of Distribution with Statistical Properties and Applications. AFIT Journal of Science and Engineering Research, 2(2), 84-103.

Sadiq, I. A., Doguwa, S. I., Yahaya, A., & Garba, J. (2023a). New Generalized Odd Fréchet-G (NGOF-G) Family of Distributions with Statistical Properties and Applications. UMYU Scientifica, 2(3), 100-107. [https://doi.org/10.56919/usci.2323.016]

Sadiq, I. A., Garba, S., Kajuru, J. Y., Usman, A., Ishaq, A. I., Zakari, Y., ... & Yahaya, A. (2024). The Odd Rayleigh-G Family of Distribution: Properties, Applications, and Performance Comparisons. FUDMA Journal of Sciences, 8(6), 514–527. [https://doi.org/10.33003/fjs-2024-0806-3011]

Sadiq, I. A., Kajuru, J. Y., Doguwa, S. I., Yahaya, G. Y., Hephzibah, A. A., Yahaya, S. S., ... & Bello, A. (2025b). Survival analysis in advanced lung cancer using Weibull survival regression model: estimation, interpretation, and clinical application. Journal of Statistical Sciences and Computational Intelligence, 1(2), 106-123.

Sadiq, I. A., Zakari, Y., Doguwa, S. I., Isiya, M., Suleiman, S. Y., Ajayi, A. H., ... & Samuel, W. (2025a). Survival Analysis of Acute Myocardial Infarction Patients Using the Kumaraswamy-Logistic Model and Kaplan-Meier Estimation. Iraqi Statisticians Journal, 15-32.

Saikh, A., & Gazi, N. H. (2021). The effect of the force of infection and treatment on the disease dynamics of an SIS epidemic model with immigrants. *Results in Control and Optimisation, 2*, 100007.

Wariri, O., Muhammad, A. K., Sowe, A., Strandmark, J., Utazi, C. E., Metcalf, C. J. E., et al. (2025). Serological survey to determine measles and rubella immunity gaps across age and geographic locations in The Gambia: A study protocol. *Global Health Action, 18*(1), 2540135.