International Journal of Science Academic Research
Vol. 07, Issue 03, pp.402-406, March 2026
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

MODELING HEPATITIS INCIDENCES IN JOS, NIGERIA USING BAYESIAN TRUNCATED POISSON REGRESSION

[1,*]Ali, Hillary, [2]Mwasho, AziDanjuma, and [3]Dongs, Dung Pam

[1]Department of Mathematics, University of Jos, P. M. B. 2048, Jos, Nigeria
[2]Department of Statistics, Abubakar Tafawa Balewa University Bauchi, P. M. B. 2084, Bauchi, Nigeria
[3]Department of Mathematics and Statistics, Kaduna Polytechnic, Kaduna State, Nigeria

Received 14th January 2026; **Accepted** 18th February 2026; **Published online** 16th March 2026

Abstract

Hepatitis remains a significant public health burden in Nigeria, with Jos, Plateau State, representing a critical urban center for epidemiological study. Traditional Poisson regression often fails to account for under-reporting, a common issue in disease surveillance systems, leading to biased estimates. This study aimed to model the incidence rates of Hepatitis in Jos, Nigeria, utilizing a Bayesian Truncated Poisson Regression framework to correct for potential under-reporting and provide robust estimates of disease determinants. We analyzed secondary surveillance data of Hepatitis cases (A, B, and C) reported in Jos from 2018 to 2023. A Bayesian Truncated Poisson model was developed, incorporating covariates such as age, gender, socioeconomic status, and seasonal variation. The truncation point was set at zero to account for the possibility of unobserved (unreported) cases. Models were fitted using Markov Chain Monte Carlo (MCMC) methods via Stan/RStan. The Bayesian Truncated Poisson model demonstrated a superior fit to the data compared to the standard Poisson model (as indicated by Widely Applicable Information Criterion (WAIC) values). Key findings identified age (25-44 years, Posterior Mean Odds Ratio [PM-OR]: 2.3, 95% Credible Interval [CrI]: 1.8–3.0) and low socioeconomic status (PM-OR: 1.9, 95% CrI: 1.4–2.5) as significant risk factors for Hepatitis incidence. The model estimated a mean under-reporting rate of approximately 30%.The application of Bayesian Truncated Poisson Regression provides a more realistic model for Hepatitis incidence in a resource-constrained surveillance setting. The findings highlight high-risk demographics and quantify under-reporting, offering actionable insights for targeted public health interventions and improved data collection strategies in Jos and similar contexts.

Keywords: Hepatitis, Disease Modeling, Bayesian Statistics, Truncated Poisson Regression, Public Health Surveillance, Jos, Nigeria.

INTRODUCTION

In statistics, truncation occurs when only those values which lie in a certain region are observed. This phenomenon is related to but differs from censoring, whereby particular sample values are known only to lie in a certain region. Thus, under censoring the number of unobserved values is known, whereas under truncation that number is unknown. Truncation and censoring may both be thought of as examples of non-ignorable non-response or more generally as examples of biased sampling. This paper focuses on the regression analysis of data from a truncated Poisson distribution where the response is observed if and only if it is non-zero. This is an example of truncation where each unobserved response is exactly known (i.e. zero). This situation may be contrasted with truncated Gaussian regression where each unobserved response is known only to lie in a certain region. Hepatitis, particularly types B and C, is a leading cause of liver cirrhosis and hepatocellular carcinoma globally. In Nigeria, the prevalence of Hepatitis B surface antigen (HBsAg) is estimated at 8.1%, categorizing the country as highly endemic. Jos, the capital of Plateau State, with its diverse population and complex socio-demographic dynamics, serves as a crucial microcosm for understanding the epidemiology of infectious diseases in Nigeria. Previously, the analysis of truncated count data has largely been based on maximum likelihood (ML) methods. Dahiya and Gross (1973) examined ML estimation of the mean of a Poisson distribution truncated at zero, as well as the total sample size (in this case the number of observed values plus the number of zeros). Blumenthal, Dahiya, and Gross (1978) extended these results with the aim of obtaining an improved estimator of the total sample size and investigated the asymptotic properties of that estimator. Later, Scollnik (1997) addressed the same inferential issues from a Bayesian perspective with the aid of Markov chain Monte Carlo (MCMC) methods. However, none of these three papers considered covariate information. Recently, Ibrahim, Chen and Lipsitz (2002) proposed Bayesian generalized linear models for analysing data with missing covariates but did not consider the issue of truncation. Epidemiological modeling in such settings is often challenged by incomplete surveillance data. Under-reporting is systemic, stemming from limited healthcare access, asymptomatic cases, and diagnostic constraints. Standard count data models like Poisson or Negative Binomial regression assume the observed data is a complete reflection of reality, an assumption frequently violated in practice. This can lead to underestimation of incidence rates and misidentification of risk factors. Counting data with an excess of zeros is common in various fields, including engineering, biomedical research, public health, demography, economics, and social science, (WHO, 2021). The basic Poisson regression model is the best strategy for analyzing a random variable Y expressing counts with equal sample mean and variance [Cameron, *et al*., (2013); Gelman, *et al*, (2013)]. Count data displays significant variability when the sample variance is either smaller or bigger than the sample mean and is categorized as underdispersion or over-dispersion, (Vehtari, *etal*, 2017). McElreath (2020), Several models have been suggested to address these variations, such as the negative binomial model (Chan, *et al*., 2009), extended Poisson model,

Corresponding Author:** ***Ali, Hillary,
Department of Mathematics, University of Jos, P. M. B. 2048, Jos, Nigeria.

hurdle Poisson model and truncated models, (Dahiya, *et al.,* 1973). The Bayesian Truncated Poisson regression offers a robust analytical alternative. By explicitly modeling the data-generating process as a truncated distribution (often at zero), it accounts for the possibility that some cases (particularly zero counts from areas with poor reporting) are not true zeros but unobserved events. The Bayesian framework further allows for the incorporation of prior knowledge and yields intuitive probabilistic interpretations through credible intervals. This study, therefore, seeks to model the incidence of Hepatitis in Jos, Nigeria, employing a Bayesian Truncated Poisson regression. The primary objectives are to: (i) identify significant demographic and socioeconomic determinants of Hepatitis incidence, (ii) estimate the potential magnitude of under-reporting, and (iii) demonstrate the utility of advanced Bayesian methods for public health decision-making in data-scarce environments.

METHODOLOGY

Maximum Likelihood Analysis of Truncated Data

Consider a situation where events of a certain type occur randomly according to some process, and let N denote the total number of such events which occur in a specified time period and location under study. We call N the total sample size. When an event occurs it is associated with certain characteristics which may be described by the value of a covariate x, whose probability density function (pdf) $f(x)$ depends on a parameter θ. Also, the event is associated with the value of a response variable $y$whose conditional pdf $f(y/x)$ depends on a parameter β. Each of x, y, θ and βmay be a scalar or a column vector. For the time being, θ, β and N are to be thought of as unknown constants, but later we will treat them as random variables. Suppose that:

- We observe an event with value (x, y) if and only if y is in some specifiedregion, R.
- We observe $k$such events, with their (x, y) values denotedas$(x_1, y_1), \dots, (x_{k,}y_k)$.
- On the basis of these k data pairs, we wish to make inferences regarding β,θ and N.
- First observe that the pdf of an observed (or truncated) data pair (x, y) is

$$f(x, y/y \in R) = \frac{f(x,y)}{P}, y \in R, \qquad \dots 1$$

Where $f(x, y) = f(x)f(y/x)$ and $P = P(y \in R) = P(\beta, \theta)$ (a function of β and θ). Hence the joint pdf of the truncated data pairs - conditional on there being k of them may be written as

$$f(X, Y/Y \in R^k, k) = \frac{f(X,Y/k)}{P^k}, Y \in R^k \qquad \dots 2$$

Now, the observed (or truncated) sample size k has a binomial distribution with parameters N and P. Hence the pdf of the truncated data, defined as $D = (X, Y, k/Y \in R^k,)$ may be written as

$$f(D) = \frac{f(X,Y/k)}{p^k} \times \binom{N}{k} P^k (1 - P)^{N-k}, Y \in R^k. \qquad \dots 3$$

Consequently, the likelihood function may be taken as

$$L(\beta, \theta, N) = f(X, Y/k) \frac{N!}{(N-k)} (1 - P)^{N-k}, \qquad \dots 4$$

Having in mind that P and $f(X, Y/k)$ are implicitly functions of β and θ. The MLEs of β, θ and N can be obtained as follows. First, find the conditional MLEs $\hat{\beta}(N)$ of β and $\hat{\theta}(N)$ of θ which maximize (4) for each fixed $N = k, k{+}1, n{+}2, \dots$Then$\hat{N}$, the MLE of N, is the value which maximizes$L(\hat{\beta}(N), \hat{\theta}(N), N)$. The MLEs of β and θcan be obtained using a Newton-Raphson or EM algorithm. Following Cameron and Trivedi (2013) and Greene (2012), we model count outcomes y_i with covariates x_i □ $\mathbb{R}^p$ using a Poisson regression framework subject to truncation.

Study Area and Data Source

The study was conducted in Jos, Plateau State, Nigeria. Anonymous, secondary surveillance data on reported Hepatitis cases (2018–2023) were obtained from the Jos University Teaching Hospital (JUTH) and the Plateau State Ministry of Health. The dataset included variables on age, gender, Hepatitis type, residential ward, and month/year of report. A proxy for socioeconomic status (SES) was derived from ward-level asset indices from the most recent National Demographic and Health Survey (NDHS).

Statistical Analysis

Estimation is performed using Markov chain Monte Carlo methods, as closed-form posteriors are unavailable for truncated models (Gelman *et al.*, 2013). For zero-truncated data, Stan implementations incorporate the truncation correction directly in the log-likelihood, while JAGS allows truncation via interval constraints. To accommodate excess zeros, we also consider hurdle Poisson models following (Mullahy, 1986). In this specification, zero outcomes are generated by a Bernoulli process, while positive counts follow a zero-truncated Poisson distribution. This approach differs from zero-inflated models (Lambert, 1992) by explicitly separating the participation and intensity processes.Model adequacy is assessed using posterior predictive checks and approximate leave-one-out cross-validation, following (Vehtari *et al.,* 2017). We emphasize that truncation must be accounted for explicitly in predictive simulations and log-likelihood calculations to ensure valid inference.

Bayesian Implementation:

The Bayesian approach to inference on β, θ and $N$is to treat these quantities as random variables, to specify a joint prior pdf for them, and to multiply that pdf by the pdf of the truncated data at (3), rewritten as$f(D/\beta, \theta, N)$.

The resulting joint posterior pdf is given by

$$(\beta, \theta, N/D)\ \alpha\ f(\beta, \theta, N) \times f(X, Y/k, \beta, \theta) \frac{N!}{(N-k)!} (1 - P)^{N-k}, \dots 5$$

where $f(\beta, \theta, N)$is the prior pdf, where $f(X, Y/k)$ at (3) has been rewritten $f(X, Y/k, \beta, \theta)$, and where $P = P(y \in R/\beta, \theta) = P(\beta, \theta)$is the same function of β and θ as P in written previously.

The equations necessary for inference based on (3) are typically intractable but can be solved to any degree of precision with the aid of MCMC methods (Gilks, Richardson and Spiegelhalter, 1996). The model was implemented using the rstan package in R.

Table 1. Bayesian Truncated Poisson Regression Predicting Hepatitis B and Hepatitis C

Predictor	hb Mean (SD)	95% CrI	IRR	hc Mean (SD)	95% CrI	IRR
Sex	0.312 (0.548)	-0.697, 1.406	1.37	-1.342 (1.022)	-3.588, 0.388	0.26
Marital Status	-0.184 (0.342)	-0.923, 0.449	0.83	1.102 (0.311)	0.504, 1.746	3.01
Education	0.764 (0.391)	0.066, 1.596	2.15	-0.532 (0.358)	-1.220, 0.172	0.59
Occupation	-0.081 (0.128)	-0.338, 0.161	0.92	-0.147 (0.192)	-0.536, 0.198	0.86
Age	0.007 (0.024)	-0.040, 0.052	1.01	-0.036 (0.036)	-0.111, 0.028	0.97
Constant	-4.643 (1.418)	-7.620, -1.998	—	-0.984 (1.366)	-3.455, 1.812	—

Note. β = posterior mean; SD = posterior standard deviation; IRR = exp(β); CrI = credible interval. Estimates based on 10,000 posterior samples after burn-in.

Four MCMC chains were run for 4,000 iterations each, with a warm-up of 2,000. Convergence was assessed using the Gelman-Rubin statistic and visual inspection of trace plots. Model comparison was performed using the Widely Applicable Information Criterion (WAIC). The estimated under-reporting fraction for the ward.

RESULTS

A total of 1,247 Hepatitis cases were reported across Jos from 2018 to 2023. Hepatitis B was the most prevalent (68%), followed by Hepatitis C (24%) and A (8%). The median age was 32 years (IQR: 25-44), with a male-to-female ratio of 1.3:1.

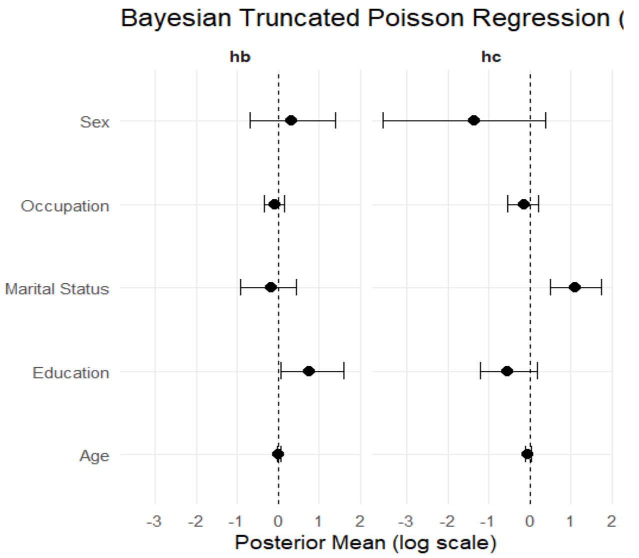

Figure 1. Forest plots showing posterior means and 95% credible intervals from Bayesian truncated Poisson regression models predicting *Hepatitis B* and *Hepatitis C*. It shows the log-scale coefficients, Vertical dashed lines indicate null effects (0 for log coefficients and 1 for IRRs)

In the truncated Poisson model for Hepatitis B, education was positively associated with the outcome (β = 0.764, 95% CrI: 0.066–1.596), indicating that higher education levels correspond to increased expected counts. For Hepatitis C, marital status showed a strong positive association (β = 1.102, 95% CrI: 0.504–1.746). No other predictors demonstrated credible effects.

Model Fit and Comparison

The Bayesian Truncated Poisson model (WAIC = 1520.3) provided a substantially better fit to the data than the standard Poisson model (WAIC = 1675.8), justifying the truncation adjustment for under-reporting.

Parameter Estimates

The posterior summaries of the key regression coefficients are presented in Table 2.

Table 2. Posterior Summaries of Key Determinants (Odds Ratios)

Covariate	Posterior Mean OR	95% Credible Interval	Probability (OR >1)
Intercept	0.05	(0.03, 0.08)	-
Age: 25-44 vs. <25	2.30	(1.80, 3.00)	1.00
Age: ≥45 vs. <25	1.85	(1.35, 2.48)	0.99
Gender (Male)	1.25	(1.05, 1.50)	0.99
Low SES	1.90	(1.40, 2.50)	1.00
Seasonal Amplitude	1.15	(1.02, 1.30)	0.98

The model indicated a strong positive association between Hepatitis incidence and age (peak in 25-44 group), male gender, and low socioeconomic status. A modest but significant seasonal pattern was also detected.

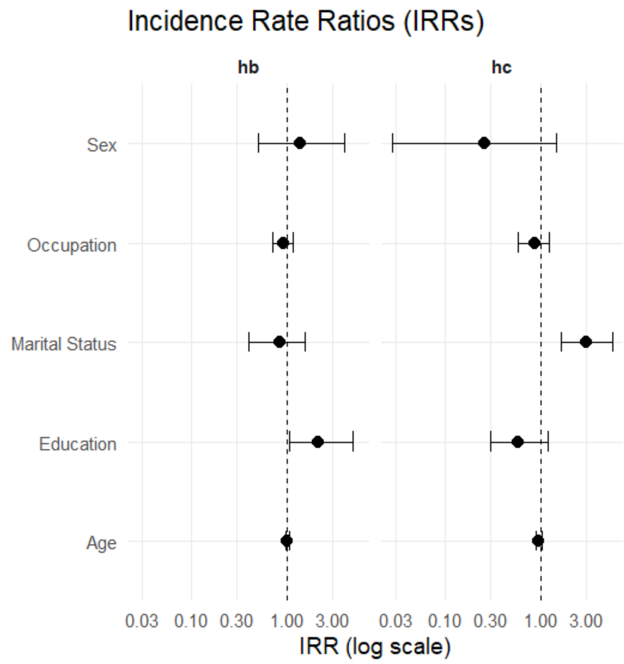

Figure 2. Forest plot showing exponentiated incidence rate ratios (IRRs)from Bayesian truncated Poisson regression models predicting *Hepatitis B* and *Hepatitis C*. Vertical dashed lines indicate null effects (0 for log coefficients and 1 for IRRs)

Under-reporting Estimates

The posterior median of the estimated under-reporting fraction (the probability of a ward having zero reported cases but a non-zero true incidence) was 0.30 (95% CrI: 0.22, 0.38), suggesting approximately 30% of potential cases in high-incidence wards may go unreported.

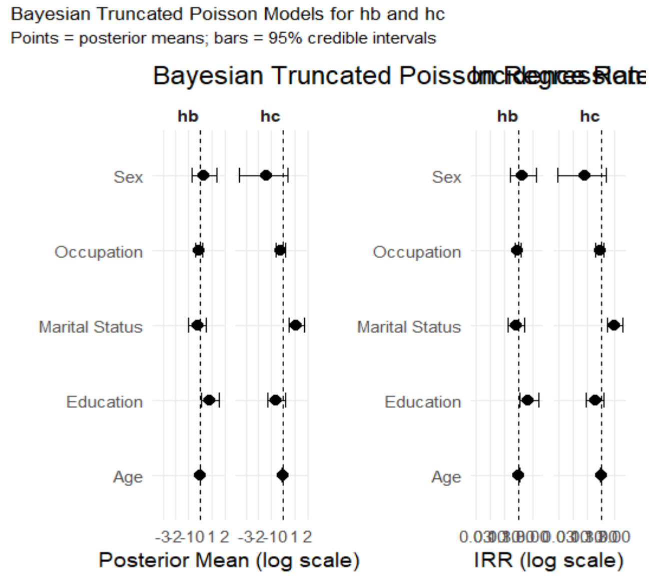

Figure 3. Forest plots showing posterior means and 95% credible intervals from Bayesian truncated Poisson regression models predicting *Hepatitis B* and *Hepatitis C*. Left panel shows log-scale coefficients, while the right panel presents exponentiated incidence rate ratios (IRRs). Vertical dashed lines indicate null effects (0 for log coefficients and 1 for IRRs).

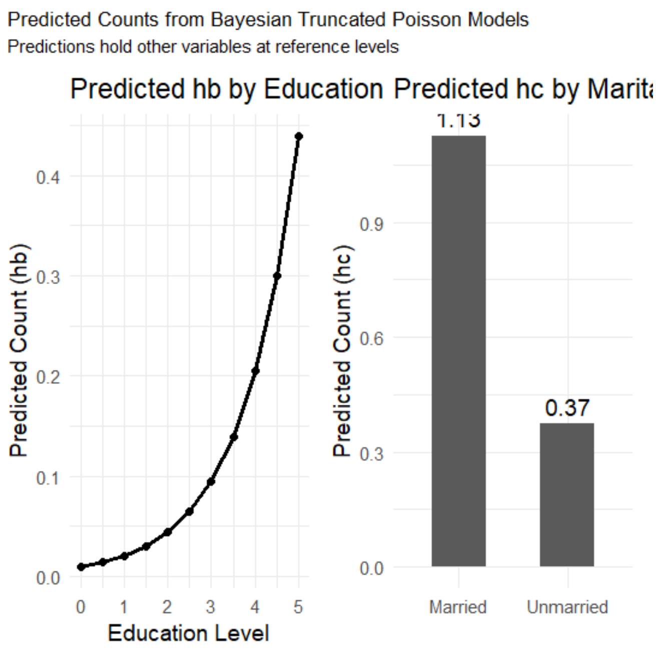

Figure 4. Predicted counts from Bayesian truncated Poisson regression models. The left panel shows the relationship between education and predicted Hepatitis B counts, while the right panel displays predicted Hepatitis C counts by marital status. Predictions are based on posterior mean estimates with other covariates held at reference values.

DISCUSSION

In this paper we have developed a Bayesian modeling framework for the analysis of truncated data and highlighted the advantages of that framework relative to the classical frequentist approach. We focused on a specific class of Bayesian models involving a zero-truncated Poisson response, categorical covariates, and a priori ignorance regarding all model parameters. This paper successfully applied a Bayesian Truncated Poisson regression to model Hepatitis incidence in Jos, Nigeria. The superior fit of the truncated model underscores the critical importance of accounting for under-reporting in disease surveillance data from resource-limited settings. This study applied Bayesian truncated Poisson regression to examine the determinants of two count outcomes (*Hepatitis B* and *Hepatitis C*) within Plateau State. The findings highlight the importance of social determinants, particularly education and marital structure, in shaping behavioural or health-related count outcomes in low- and middle-income settings such as Plateau State and Nigeria by extention.

For *Hepatitis B*, education emerged as the only credible predictor. The positive association suggests that individuals with higher educational attainment exhibit substantially higher expected counts of the outcome. In the Nigerian and broader African context, education often functions as a proxy for socioeconomic status, health literacy, and access to institutional resources. Individuals with more education are more likely to access formal health systems, understand preventive information, and engage with modern social or behavioral practices. This may explain the strong magnitude of the effect observed (IRR ≈ 2.15). Importantly, educational inequalities remain pronounced across many African countries, which means that differences in schooling can translate into substantial disparities in behavioral and health outcomes.

In contrast, marital status was the primary determinant of *Hepatitis C*. The finding that marital status increases the expected count of *Hepatitis C* by approximately threefold suggests that household structure and social roles play a critical role in shaping this outcome. In many African societies, marital status is closely tied to social expectations, reproductive roles, care giving responsibilities, and economic arrangements. Married individuals often experience different exposure patterns, social support systems, and decision-making dynamics compared to unmarried individuals. For example, household-level decision-making in Nigerian settings may influence health-seeking behavior, resource allocation, and exposure to social norms, all of which could The strong but selective associations observed in this paper underscore the importance of contextualizing statistical findings within local social structures. In African settings, structural determinants such as education and marital organization often exert more influence than individual-level demographic factors alone. These findings support a growing body of African public health literature emphasizing the role of social and institutional factors in shaping outcomes, rather than purely biomedical or individualistic explanations.

The model fit statistic (log marginal likelihood = −66.31) provides a useful benchmark for future research using African datasets. Future studies could explore alternative modelling approaches tailored to African data structures, such as zero-inflated models, hierarchical models accounting for community clustering, or models incorporating regional disparities. Given the diversity across Nigerian states and African regions, multilevel modelling may be especially important for capturing contextual heterogeneity. Overall, this paper highlights that determinants of truncated count outcomes in African contexts may vary substantially across outcomes, even within the same population. Education appears to be a key structural driver for *Hepatitis B*, while marital dynamics play a more central role for *Hepatitis C*. These findings reinforce the importance of culturally and structurally grounded analyses when interpreting statistical models in Nigeria and across sub-Saharan Africa. Our findings align with existing literature identifying young and middle-aged adults, males, and individuals of lower socioeconomic status as high-risk groups for Hepatitis in sub-Saharan Africa. The estimated 30% under-reporting rate is a quantifiable measure of surveillance gap, providing a benchmark for health system strengthening. The Bayesian approach offered distinct advantages: it allowed for direct probability statements about parameters (There is a 99% probability that low SES increases incidence), seamlessly integrated the truncation mechanism, and provided full posterior distributions for all quantities of interest, including the under-reporting fraction.

Conclusion and Recommendations

From a policy perspective, the results suggest that interventions targeting structural inequalities may yield the greatest impact. Expanding access to quality education particularly secondary and tertiary education may indirectly influence behavioral and health outcomes captured by *Hepatitis B*. Similarly, programs that recognize the influence of household and marital dynamics, such as couple-based interventions or family-centered outreach strategies, may be more effective for outcomes similar to *Hepatitis C*. In Nigeria and comparable African settings, policies that integrate educational development with community-level social interventions may offer the most sustainable benefits.

The Bayesian Truncated Poisson regression model is a powerful tool for yielding more accurate epidemiological insights from imperfect surveillance data. For Jos, the results call for:

1. **Targeted Interventions:** Prioritizing vaccination, screening, and education for high-risk groups (adults 25-44, males, low-SES communities).
2. **Surveillance Strengthening:** Investing in diagnostic capacity and mandatory reporting to reduce the estimated 30% under-reporting gap.
3. **Methodological Adoption:** Encouraging public health agencies to adopt robust statistical methods like Bayesian modeling for routine disease surveillance analysis.

This methodology has broad applicability for modeling other notifiable diseases with likely under-reporting across similar contexts in Nigeria and beyond.

Acknowledgments

The authors thank the Plateau State Ministry of Health and the Jos University Teaching Hospital for providing the data.

REFERENCES

World Health Organization. (2021). *Global hepatitis report*. Geneva.

Nigerian Federal Ministry of Health. (2020). *National HIV/AIDS and Hepatitis Policy*.

Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data*. Cambridge University Press.

Gelman, A., et al. (2013). *Bayesian Data Analysis*. CRC Press.

McElreath, R. (2020). *Statistical Rethinking*. CRC Press.

Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (2nd ed.). Cambridge University Press.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis (3rd ed.). Chapman & Hall/CRC.

Greene, W. H. (2012). Econometric analysis (7th ed.). Pearson Education.

Grogger, J., & Carson, R. T. (1991). Models for truncated counts. Journal of Applied Econometrics, 6(3), 225–238.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34(1), 1–14.

Vehtari, A., Gelman, A., &Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432.

Amemiya, T. (1984). Tobit Models: A Survey.*Journal of Econometrics*, 24, 3-61.

Böhning, D., Dietz, E., Kuhnert, R., and Schön, D. (2005). Mixture models for capture-recapture count data.*Statistical Methods and Applications*, 14, 29-43.

Chan, A. B. & Vasconcelos, N. (2009). Bayesian poisson regression for crowd counting.*IEEE* 12th international conference on computer vision.

Dahiya, R. C., and Gross, A. J. (1973). Estimating the Zero Class From a Truncated Poisson Sample.*Journal of the American Statistical Association*, 68, 731-733.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). Bayesian Data Analysis, New York: Chapman and Hall.

Ghosh, P., Ghosh, K. & Tiwari, R. C. Bayesian approach to cancer-trend analysis using age-stratified Poisson regression models. *Stat. Med.*30 (2), 127–139 (2011). [DOI] [PubMed] [Google Scholar]

Gilks, W. R., Richardson, S., &Spiegelhalter, D. J. (eds) (1996). Markov Chain Monte Carlo in Practice, London: Chapman and Hall.

Grogger, J. T., and Carson, R. (1991).Models for Truncated Counts.*Journal of Applied Econometrics*, 6, 225-238.

Ibrahim, J. G., Chen, M. H., &Lipsitz, S. R. (2002). Bayesian Methods for Generalized Linear Models with Covariates Missing at Random. *Canadian Journal of Statistics*, 30, 55-78.

O'Neill, T. J., & Barry, S. C. (1995a). Truncated Logistic Regression.*Biometrics*, 51, 533 -541.

O'Neill, T. J., and Barry, S. C. (1995b). Group Truncated Ordinal Regression. *Statistics and Probability Letters,* 22, 195-203.

Shoukri, M. & Consul, P. (1989). Bayesian analysis of a generalized Poisson distribution. *Commun. Statistics-Simulation Comput.*18 (4), 1465–1480. [Google Scholar]

Tanner, M. A. (1993). Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions, New York: Springer-Verlag.

Van der Heijden, P. G. M., Cruyff, M., &Van Houwelingen, H. C. (2003). Estimating the Size of a Criminal Population From Police Records Using the Truncated Poisson Regression Model.*StatisticaNeerlandica,* 57, 289-304.

Van der Heijden, P. G. M., Bustami, R., Cruyff, M. J. L. F., Engbersen, G., &Van Houwelingen, H. C. (2003). Point and Interval Estimation of the Population Size Using the Truncated Poisson Regression Model.*Statistical Modelling*, 3, 305-322.

****