FULafia Journal of Science & Technology
ISSN (print) 2449-0954 \ ISSN (online) 2636-4972
https://lafiascijournals.org.ng/index.php/fjst
Sponsored By TERTIARY EDUCATION TRUST FUND
Manuscript => Received: Jan., 2026; Accepted: June, 2026; Published: June, 2026 https://doi.org/10.62050/fjst2026.v10n2.731

MULTILEVEL COUNT REGRESSION ANALYSIS OF FACTORS AFFECTING ANTENATAL CARE VISITS IN NIGERIA

Musa Ganaka Kubi[1], Tobias Chukwudi Ossai[2], E. O. Ben[3], Rilwan Bala Muhammad[4] and Y. B. Usman[5]
[1]*Department of Public Health, Abubakar Tafawa-Balewa University, Bauchi, Nigeria*
[2]*Department of Statistics and Data Analytics, Nasarawa State University, Keffi, Nigeria*
[3]*Department of Statistics, Abubakar Tafawa-Balewa University, Bauchi, Nigeria*
[4]*Department of Epidemiology and Biostatistics, Federal University of Health Science, Azare, Nigeria*
[5]*Department of Mathematics and Statistics, Federal Polytechnic Idah, Nigeria*
*Corresponding email: ganakamusak@gmail.com

ABSTRACT
Adequate antenatal care (ANC) is essential for maternal and neonatal health, yet utilization in Nigeria remains suboptimal. This study aimed to determine the factors influencing number of ANC visits using multilevel count data regression model. Data from 12,719 women in the 2024 Nigerian Demographic and Health Survey data were analyzed. A multilevel Zero-Inflated Negative Binomial (ZINB) model was selected over competing count models based on its superior ability to handle the data's overdispersion, substantial zero-inflation (28 %) and clustered structure. Both count (frequency) and zero-inflation (non-visit) components were estimated.In the count component, higher maternal age (25–39: AIRR =1.053; 40-49: AIRR =1.073), maternal secondary (AIRR =1.069) and higher education (AIRR =1.113), husband's secondary (AIRR =1.077) and higher education (AIRR =1.113), wealthier households (middle: AIRR =1.054; rich: AIRR =1.116), media exposure (AIRR =1.054), employment (AIRR = 1.035), insurance (AIRR =1.080), shorter distance to health facilities (AIRR =1.040), and prior pregnancy termination (AIRR =1.036) were associated with increased ANC visits. In the zero-inflation component, maternal and husband education (AORs =0.530–0.205), wealth (AOR middle =0.731; rich =0.355), employment (AOR =0.641) and shorter facility distance (AOR =0.763) reduced the likelihood of non-attendance, whereas higher parity (2-4: AOR =1.408; ≥5: AOR =1.614) and husband-dominated decision-making (AOR =1.297) increased it. ANC utilization is shaped by demographic, socioeconomic, educational, and access-related factors, with pronounced regional and parity disparities. Targeted interventions are needed to improve equitable ANC coverage.

Keywords: Antenatal care visit, Zero-inflated negative binomial, Multilevel count data modeling

INTRODUCTION
Antenatal care (ANC) constitutes a fundamental component of maternal healthcare, designed to optimize pregnancy outcomes through routine monitoring, health education, and early detection of complications. The World Health Organization (WHO) has progressively strengthened its recommendations, moving from a minimum of four visits to a standard of eight contacts for a positive pregnancy experience (World Health Organization, 2016). Despite these global standards, the utilization of ANC services remains a critical public health challenge in many low- and middle-income countries (LMICs). In Nigeria, the situation presents particular concern, with the 2024 Nigeria Demographic and Health Survey revealing that 37 % of pregnant women received no ANC services (based on an overall coverage rate of 63 %), while significant disparities persist across geographical regions and socioeconomic strata (National Population Commission [NPC] & ICF, 2024). Previous research has consistently identified regional variations, with women in southern Nigeria demonstrating higher utilization rates compared to their northern counterparts, alongside strong associations between ANC attendance and factors such as maternal age, urban residence, and wealth quintile (Yusuf & Ugalahi, 2015; Yusuf *et al*., 2018).
The analytical complexity of ANC utilization data stems from its inherent nature as count data, which frequently violates the fundamental assumptions of conventional statistical models. ANC visit data typically exhibits both over-dispersions, where the variance exceeds the mean, and a substantial proportion of zero counts (Yusuf *et al*., 2018; Yadav *et al*., 2021). The application of standard Poisson regression, which assumes equality of mean and variance (equi-dispersion), to such data yields biased parameter estimates and unreliable inferences (Yusuf & Ugalahi, 2015). Furthermore, the excess zeros often arise from a two-stage decision process: the initial choice to seek care, followed by decisions regarding visit frequency, a structural complexity that standard models cannot adequately capture (Bhowmik *et al*., 2020).
To navigate these methodological challenges, researchers have transitioned from standard Poisson regression toward more sophisticated count frameworks that account for distributional anomalies. Specifically,

Negative Binomial (NB) and Generalized Poisson (GP) models are frequently used to address over-dispersion (Yusuf & Ugalahi, 2015), while Zero-Inflated (ZIP, ZINB) and Hurdle models address the statistical bias introduced by excess zeros (Bhowmik *et al*., 2020). These approaches represent a critical conceptual shift: zero-inflated (ZI) models distinguish between 'structural' and 'sampling' zeros, whereas hurdle models assume a two-stage process where zeros and positive counts originate from distinct mechanisms. However, while these frameworks address the numerical distribution of ANC visits, they often overlook the hierarchical nature of health survey data a gap this study addresses by integrating these count models into a multilevel structure.

Empirical evidence across diverse contexts demonstrates considerable variation in optimal model selection. In Nigeria, Yusuf and Ugalahi (2015) found the Generalized Poisson model superior for handling over-dispersion in ANC data, while Yusuf *et al*. (2018) subsequently identified the Zero-Inflated Negative Binomial (ZINB) model as optimal when accounting for excess zeros. This pattern of context-dependent model performance extends beyond Nigeria. In India, Yadav *et al*. (2021) identified ZINB as the best-fitting model, whereas in Bangladesh, Ahmed and Mallick (2019) and Bhowmik et al. (2020) found Hurdle Negative Binomial specifications to be superior. Similar inconsistencies appear in Ethiopia, where Zemene and Birhanu (2022) and Woldeyohannes *et al*. (2023) identified ZIP as optimal, while Bekalo and Kebede (2021) and Suleman *et al*. (2023) found Hurdle models preferable.

A critical methodological advancement in this domain involves recognizing the hierarchical structure of Demographic and Health Survey (DHS) data, where women are nested within clusters and regions. Standard count models fail to account for this clustering, potentially leading to underestimated standard errors and invalid inferences. Recent studies have demonstrated the superiority of multilevel count approaches. Bhowmik *et al*. (2020) showed that incorporating cluster-specific random effects significantly improved model fit in Bangladesh, while Suleman *et al*. (2023) established the multilevel Hurdle Negative Binomial model as optimal for Ethiopian data, with regional variations accounting for 11 % of the variance in ANC utilization. Similarly, Gebeyehu *et al*. (2022) successfully applied multilevel negative binomial regression across Sub-Saharan Africa, and Negash *et al*. (2025) utilized multilevel Poisson modeling in Senegal, confirming the importance of accounting for hierarchical data structures.

This literature review reveals two significant gaps in the Nigerian context. First, existing studies have produced conflicting recommendations regarding optimal count models, with Yusuf and Ugalahi (2015) advocating for Generalized Poisson while Yusuf *et al*. (2018) favored the ZINB framework. Second, and more critically, no previous study in Nigeria has incorporated multilevel frameworks into count model comparisons, despite evidence that ignoring hierarchical structures leads to underestimated standard errors and invalid inferences (Bhowmik *et al*., 2020). Beyond mere statistical fit, a multilevel ZINB approach is specifically required to disentangle the double hurdle of ANC utilization in Nigeria: it accounts for the structural zeros women are entirely excluded from care due to systemic barriers, while simultaneously capturing how regional and community-level clustering influences the frequency of visits for those who do enter the system. This represents a substantial methodological limitation in current Nigerian research, particularly given that regional disparities are often rooted in community-level norms rather than just individual-level factors.

This study therefore aims to determine the best fitting multilevel count model for Nigerian ANC data and also to identify the individual and community-level factors associated with ANC utilization, thereby informing targeted interventions and contributing to methodological advancements in analyzing complex health survey data.

MATERIALS AND METHODS

Source of Data

The data utilized in this study were obtained from the 2024 Nigeria Demographic and Health Survey (NDHS), a nationally representative survey conducted by the National Population Commission (NPC) in collaboration with the DHS Program. The NDHS employs a stratified, multistage cluster sampling design to collect detailed information on maternal and child health, fertility, household characteristics, and healthcare utilization across all geopolitical zones of Nigeria. The survey provides high-quality, population-based data that are widely used for research, policy formulation, and program evaluation. For this study, relevant variables related to number of antenatal care visit, maternal socio-demographic characteristics, and cluster-level factors were extracted and prepared for analysis.

Variables Description

The variables of the study include maternal, household and community-level characteristics. The maternal age was categorized as 15–24, 25–39, and 40–49 years, while parity was grouped as first, 2–4, or 5+ births. Maternal and husband education levels were recorded as none, primary, secondary, or higher. Socioeconomic and contextual factors included wealth index (poor, middle, rich), place of residence (urban, rural), and region (NW, NE, NC, SE, SS, SW). Other variables comprised media exposure, religion, working status, health insurance coverage, decision-making on maternal health, distance to health facilities, birth order, birth type (single or multiple), and history of terminated pregnancy. These variables were included to examine their association with number of antenatal care utilization among Nigerian women.

Techniques for Data Analysis

This study modeled the number of ANC visits among Nigerian women using multilevel count regression models that account for the hierarchical structure of the Nigerian Demographic and Health Survey Data (NDHS), where women (level 1) are nested within clusters (level 2).

Let y_{ij} denote the number of ANC visits for woman i in community j, with fixed-effect covariates X_{ij} capturing individual- and household-level factors and a random cluster-level effect b_j. To account for unobserved contextual heterogeneity, all models include a random intercept, let assumed $b_j \sim N(0, \sigma_b^2)$ and let $\mu_{ij} = \mathrm{E}(y_{ij}|X_{ij}, b_j)$ denotes the conditional mean. The general linear predictor for the multilevel models is:

$$\log(\mu_{ij}) = X_{ij}\beta + b_j \qquad (1)$$

All multilevel count models were fitted using the glmmTMB package in R, which implements maximum likelihood estimation via Laplace approximation and automatic differentiation to efficiently handle mixed-effects count models (Brooks *et al.*, 2017).

Multilevel Poisson Regression Model

The multilevel Poisson model serves as the baseline count model and assumes that the conditional variance equals the mean. It is appropriate when overdispersion is minimal.

$$y_{ij}|b_j \sim Poisson(\mu_{ij}),\ \log(\mu_{ij}) = X_{ij}\beta + b_j \qquad (2)$$

Although widely used, the Poisson model can underestimate variability when overdispersion is present (Cameron & Trivedi, 1998). Therefore, more flexible alternatives were employed.

Multilevel Negative Binomial Regression (Linear Overdispersion)

The linear overdispersion (NB1) model relaxes the strict Poisson variance assumption by allowing variance to grow linearly with the mean appropriate for mild overdispersion.

$$y_{ij}|b_j \sim NB1(\mu_{ij}, \delta),$$
$$\log(\mu_{ij}) = X_{ij}\beta + b_j,$$
$$Var(y_{ij}|b_j) = \mu_{ij} + \delta\mu_{ij},\ \delta > 0. \qquad (3)$$

Where δ is the over-dispersion parameter of NB1. This formulation captures under- or over-dispersion through δ (Hilbe, 2011).

Multilevel Negative Binomial Regression (Quadratic Overdispersion)

The quadratic overdispersion (NB2) model is preferred when overdispersion increases more sharply. It assumes that the variance grows quadratically with the mean, a common phenomenon in health service utilization.

$$y_{ij}|b_j \sim NB2(\mu_{ij}, \alpha),$$
$$Var(y_{ij}|b_j) = \mu_{ij} + \alpha\mu_{ij}^2,\ \alpha > 0. \qquad (4)$$

Where α is the over-dispersion parameter of NB2. The NB2 corresponds to a Poisson-gamma mixture and is widely recommended for count data (Hilbe, 2011).

Multilevel Zero Inflated Poisson (ZIP) Model

Number of ANC visits data often exhibit excess zeros, representing women who received no ANC care. The ZIP model distinguishes structural zeros from sampling zeros through a two-component mixture (Lambert, 1992) as follows:

Zero Inflation Component

$$p(z_{ij} = 1) = \pi_{ij}, logit(\pi_{ij}) = w_{ij}\gamma + \mu_j, \qquad (5)$$

Count Component

$$y_{ij}|(z_{ij} = 0, b_j) \sim Poisson(\mu_{ij}), \log(\mu_{ij}) = X_{ij}\beta + b_j \qquad (6)$$

Mixture Probability

$$p(y_{ij} = 0) = \pi_{ij} + (1 - \pi_{ij})e^{-\mu_{ij}} \qquad (7)$$

Where z_{ij} is the latent (unobserved) zero inflation indicator, π_{ij} is the probability of structural zeros, w_{ij} is the vector of covariate predicting zero inflation and γ measures the effect of zero inflation covariates on the probability of structural zeros.

Multilevel Zero Inflated Negative Binomial (ZINB) Model

The ZINB model generalizes ZIP by simultaneously addressing overdispersion and excess zeros, making it appropriate when variability and zero inflation co-exist.

Zero Inflation Component

$$logit(\pi_{ij}) = w_{ij}\gamma + \mu_j \qquad (8)$$

Count Component

$$y_{ij}|(z_{ij} = 0, b_j) \sim NB2(\mu_{ij}, \alpha),\ \log(\mu_{ij}) = X_{ij}\beta + b_j \qquad (9)$$

Mixture Probability

$$p(y_{ij} = 0) = \pi_{ij} + (1 - \pi_{ij})\overset{p}{}_{NB2}(0|\mu_{ij}, \alpha) \qquad (10)$$

ZINB is particularly suitable for health interventions with both non-users and variable intensity of use.

Multilevel Hurdle Poisson (HP) Model

The hurdle model separates the ANC process into (i) the decision to initiate care and (ii) the intensity of use among attenders. Unlike ZIP, zeros arise only from the binary decision stage (Mullahy, 1986).

Binary Hurdle

$$d_{ij} = 1(y_{ij} > 0),\ logit(\phi_{ij}) = w_{ij}\gamma + \mu_j \qquad (11)$$

Positive Count (Zero Truncated Poisson)

$$p(y_{ij} = y|y > 0, b_j) = \frac{\mu_{ij}^{y}e^{-\mu_{ij}}/y!}{1 - e^{-\mu_{ij}}},\ \log(\mu_{ij}) = X_{ij}\beta + b_j \qquad (12)$$

Where d_{ij} is an indicator that the count is positive and ϕ_{ij} is the probability of a positive count.

Multilevel Hurdle Negative Binomial (HNB) Model

This variant extends the hurdle framework by allowing overdispersion in the positive ANC visit counts.

Binary Component

$$logit(\phi_{ij}) = w_{ij}\gamma + \mu_j \qquad (13)$$

Count Component (Zero Truncated NB2)

$$p(y_{ij} = y | y > 0, b_j) = \frac{P_{NB2}(y|\mu_{ij},\alpha)}{1-P_{NB2}(0|\mu_{ij},\alpha)}, \ \log(\mu_{ij}) = \ X_{ij}\beta + b_j \quad (14)$$

Multilevel Generalized Poisson (GP) Model

The Generalized Poisson model (Consul & Jain, 1973) flexibly accommodates both overdispersion and under dispersion.

$$y_{ij} \sim GP(\lambda_{ij}, \theta), -1 < \theta < 1$$

$$P(Y = y) = \frac{\lambda_{ij}(\lambda_{ij}+\theta y)^{y-1} e^{-(\lambda_{ij}+\theta y)}}{y!} \qquad (15)$$

$$\log(\lambda_{ij}) = X_{ij}\beta + b_j.$$

Mean and Variance

$$E(y_{ij}) = \frac{\lambda_{ij}}{1-\theta}, \qquad Var(y_{ij}) = \frac{\lambda_{ij}}{(1-\theta)^3}$$

Where θ is the dispersion parameter. This model is valuable when ANC counts display either unusually low or unusually high variability.

Model Estimation, Comparative Assessment and Adequacy Checks

All models were estimated using maximum likelihood with random effects integration via the Laplace approximation. Competing models were compared using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and -2Log-likelihood. The random effects (variation of effects) were estimated by intra-cluster correlation coefficient (ICC) to explain the cluster variability. It can be calculated as:

$$ICC = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\pi^2}{3}},$$

Where: σ_b^2 is the variance of the null model (Kreft *et al*., 1998; Merlo *et al*., 2006; Hox *et al*., 2017).

Over-dispersion was evaluated using the Pearson chi-square goodness-of-fit statistic from the Poisson regression model. A ratio of the Pearson chi-square statistic to its degrees of freedom greater than one was taken as evidence of over-dispersion in the data (Payne *et al*., 2018). Model adequacy was subsequently assessed using diagnostic plots, including residuals versus fitted values, a histogram of simulated test statistics compared with the observed statistic, and a Q–Q plot.

Statistical analysis was performed using the glmmTMB (Generalized Linear Mixed Models using Template Model Builder) package in R. This framework is specifically engineered for fitting generalized linear mixed models with a high degree of flexibility regarding both the conditional distribution and the zero-inflation component (Brooks *et al*., 2017). Unlike standard multilevel packages, glmmTMB utilizes the Template Model Builder (TMB) based on automatic differentiation, which allows for faster and more stable parameter estimation in complex models involving high-dimensional random effects. The package was selected for this study due to its robust ability to handle the double-hurdle nature of antenatal care data by simultaneously estimating:

The zero-inflation component: Modeling the structural zeros (non-users of ANC) via a logistic link function.

The conditional count component: Modeling the frequency of visits using Poisson, Negative Binomial, or Generalized Poisson distributions.

The random effects: Accounting for the hierarchical nesting of women within clusters/regions to prevent the underestimation of standard errors.

Parameter estimation was conducted using Maximum Likelihood based on the Laplace approximation, providing a reliable balance between computational efficiency and accuracy for non-Gaussian distributions.

RESULTS AND DISCUSSION

This section presents the findings from the statistical analyses. It begins with descriptive summaries of the study response variable, followed by results from model comparison and the fitted regression model to estimate the parameters of the predictors of numbers of ANC visit. Model diagnostics are provided to assess the adequacy, robustness, and overall performance of the selected models.

The distribution of Antenatal Care (ANC) visits among the study population (N = 12,719) in Table 1 revealed significant deviations from normality, characterized by a substantial proportion of non-utilization and a multi-modal pattern. As presented in Table 1, 27.75 % of respondents (n = 3,530) reported zero ANC visits, indicating a notable zero-inflation in the dataset.

Table 1: Frequency distribution of antenatal care visits

Number of Visits	Frequency	%	Cumulative Freq.	Cumulative %
0	3530	27.75	3530	27.75
1	328	2.58	3858	30.33
2	660	5.19	4518	35.52
3	1328	10.44	5846	45.96
4	1472	11.57	7318	57.53
5	1445	11.36	8763	68.89
6	1065	8.37	9828	77.26
7	584	4.59	10412	81.85
8^+	2307	18.14	12719	99.99

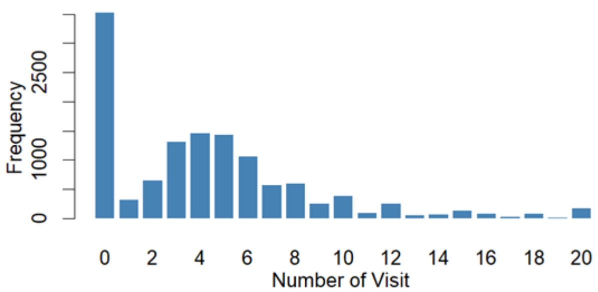

Figure 1: The number of ANC visits per woman is depicted as a bar graph

While the World Health Organization recommends a minimum of eight contacts for a positive pregnancy experience, only 18.14 % (n = 2,307) of women achieved this threshold. More than half of the sample

(57.53 %) reported four or fewer visits, with local peaks observed at 4 visits (11.57 %) and 5 visits (11.36 %). These descriptive statistics, specifically the heavy tail at zero and the over-dispersion relative to the mean provide strong empirical justification for employing zero-modified and over-dispersed multilevel count regression models rather than standard Poisson frameworks. The bar chart showing the distribution of the number of ANC visits is presented in Fig. 1.

The descriptive statistics for antenatal care (ANC) visits (N = 12,719) in Table 2 further confirmed the non-normal distribution of the data and the need for advanced count modeling. The mean number of visits was 4.41 (SD = 4.37), while the median was 4.00, reflecting a right-skewed distribution ranging from 0 to a maximum of 20 visits.

A critical finding is the presence of significant over-dispersion, as evidenced by a variance (18.81) that is substantially higher than the mean (4.41). This is quantitatively supported by a variance-to-mean ratio (VMR) of 4.27. In a standard Poisson distribution, the VMR is expected to be approximately 1.0; a ratio of 4.27 indicates severe over-dispersion, violating the equi-dispersion assumption of Poisson regression. These diagnostic metrics justify the adoption of models that account for excessive variability and clustering frameworks.

Table 2: Descriptive statistic of antenatal care visits

Parameter	Outcome
Mean (SD)	4.41(4.37)
Median ()	4.00
Q_1	0
Q_3	6
Minimum, Maximum	0
Maximum	20
Variance	18.81
Variance – to– mean ratio	4.27

Table 3: Poisson model diagnostic

Framework	Yield
Pearson χ^2	22730.35
DF	12676
χ^2/DF ratio	1.79
P-value for over-dispersion	<0.001

The diagnostic results for the initial Poisson regression model in Table 3 confirmed that the data is poorly suited for a standard Poisson framework. The Pearson chi-square statistic (χ^2=22730.35) relative to the 12,676 degrees of freedom yields a dispersion ratio of 1.79. In a well-fitted Poisson model, this ratio should ideally be close to 1.0; a value of 1.79 indicates significant over-dispersion.

This is further substantiated by the formal over-dispersion test, which yielded a p-value of < 0.001. This result leads to the rejection of the null hypothesis of equi-dispersion (where mean equals variance), confirming that the Poisson model underestimates the variability in ANC visits. Consequently, the use of more flexible models that incorporate an over-dispersion parameter, is methodologically required to ensure valid parameter estimates and standard errors.

The performance of various multilevel count regression models was evaluated using a nested modeling approach. For each model, a Null Model (intercept-only model with random effects at the cluster level) was compared against a Full Model (incorporating all individual and community-level fixed effects). This comparison specifically tests whether the inclusion of the explanatory variables significantly improves the model's predictive power over a model that only accounts for regional clustering.

As shown in Table 4, the Likelihood Ratio Test (LRT) comparing the Null and Full versions of each specification yielded statistically significant results ($p < 0.001$) across all distributions. For example, the ZINB model showed a significant improvement (LRT = 1435.4, $p < 0.001$), indicating that the covariates significantly explain the variation in ANC visits.

Table 4: Multilevel count data model comparison

	Model type	AIC	BIC	-2LL	LRT (P-value)	ICC
Poisson	Null	63955.6	63970.5	63951.6		0.832
	Full	61505.3	61766.1	61435.3	2516.3 (<0.001)	0.597
NB1	Null	59080.3	59102.6	59074.3		0.570
	Full	57175.7	57443.9	57103.7	1970.6 (<0.001)	0.240
NB2	Null	60804.9	60827.2	60798.9		0.658
	Full	58932.9	59201.1	58860.9	1938.0(<0.001)	0.355
ZIP	Null	58572.3	58594.7	58566.3		0.339
	Full	57177.0	57445.2	57105.0	1461.3 (<0.001)	0.124
ZINB	Null	58275.7	58305.5	58267.7		0.711
	Full	56906.3	57182.0	56832.3	1435.4 (<0.001)	0.390
HP	Null	59415.8	59438.1	59409.8		0.559
	Full	57816.4	58084.6	57744.4	1665.4 (<0.001)	0.217
HNB	Null	59197.6	59227.4	59189.6		0.507
	Full	57634.7	57910.4	57560.7	1628.9 (<0.001)	0.166
GP	Null	59305.2	59327.5	59299.2		0.642
	Full	57387.3	57655.5	57315.3	1983.9 (<0.001)	0.309

The substantial Intraclass Correlation Coefficients (ICC) observed in the Null models (ranging from 0.339 to 0.832) underscore that a significant proportion of the variance in ANC visits is attributable to community-level clustering. Although the ICC decreased in the Full models, it remained prominent, justifying the necessity of the multilevel framework to account for these hierarchical dependencies.

Among the eight specifications, the Multilevel Zero-Inflated Negative Binomial (ZINB) model demonstrated the superior fit, evidenced by the lowest Akaike Information Criterion (AIC = 56,906.3) and Bayesian Information Criterion (BIC = 57,182.0). This indicated that the ZINB framework most effectively handles the dual challenges of over-dispersion and structural zeros inherent in the Nigerian ANC visits dataset. While the Negative Binomial (NB1) and Zero-Inflated Poisson (ZIP) models performed competitively, the standard Poisson model exhibited the poorest fit, further confirming its inadequacy for over-dispersed health survey data.

The count component of the zero-inflated negative binomial model in Table 5 identified several factors significantly associated with the number of antenatal care (ANC) visits. Maternal age was positively associated with ANC utilization, with women aged 25–39 (AIRR = 1.053, p < 0.001) and 40–49 (AIRR = 1.073, p < 0.001) attending more visits compared to those aged 15–24. Regional differences were pronounced, as women in North Central (AIRR = 1.188, p < 0.001), South East (AIRR = 1.572, p < 0.001), South-South (AIRR = 1.668, p < 0.001) and South West (AIRR = 2.104, p < 0.001) reported higher ANC attendance than those in the North West. Maternal education was positively associated with visit frequency, particularly for secondary (AIRR = 1.069, p < 0.001) and higher education (AIRR = 1.113, p < 0.001). Similarly, husband's education influenced number of ANC visit, with secondary (AIRR = 1.077, p < 0.001) and higher education (AIRR = 1.113, p < 0.001) associated with increased visits. Socioeconomic status was significant, with women from middle- (AIRR = 1.054, p = 0.006) and rich-class households (AIRR = 1.116, p < 0.001) had more visits than those from poorer households. Birth order also influenced number of ANC visit, as third (AIRR = 0.947, p = 0.044) and fourth births (AIRR = 0.935, p = 0.018) were associated with fewer visits as compared to first birth. Rural residence was linked to lower attendance (AIRR = 0.952, p = 0.007) compared to urban, whereas media exposure (AIRR = 1.054, p < 0.001), insurance coverage (AIRR = 1.080, p = 0.003), working status (AIRR = 1.035, p = 0.004), shorter distance to health facilities (AIRR = 1.040, p = 0.003), and a history of terminated pregnancy (AIRR = 1.036, p = 0.008) were all associated with increased ANC visits. These results underscore the critical role of demographic, socioeconomic, educational, and access-related factors in determining adequate antenatal care utilization.

Table 5: Count component results of ZINB – Factors associated with number of ANC visits

Predictors	Categories	Coeff.	AIRR	95CI	P-value
Intercept	(Intercept	1.287	3.623	(3.265-4.020)	<0.001
Maternal Age	15 – 24	*Ref.*			
	25 – 39	0.052	1.053	(1.021-1.086)	<0.001
	40 – 49	0.070	1.073	(1.031-1.116)	<0.001
Parity	1	*Ref.*			
	2 – 4	-0.036	0.964	(0.919-1.012)	0.139
	5+	-0.065	0.937	(0.876-1.001)	0.055
Region	North West	*Ref.*			
	North East	0.019	1.019	(0.967-1.073)	0.477
	North Central	0.173	1.188	(1.127-1.254)	<0.001
	South East	0.452	1.572	(1.474-1.676)	<0.001
	South-South	0.511	1.668	(1.564-1.778)	<0.001
	South West	0.744	2.104	(1.985-2.231)	<0.001
Maternal Edu. Level	No Edu.	*Ref.*			
	Primary	0.027	1.027	(0.985-1.072)	0.211
	Secondary	0.066	1.069	(1.028-1.111)	<0.001
	Higher	0.107	1.113	(1.060-1.169)	<0.001
Religion	Christian	*Ref.*			
	Islam	0.013	1.013	(0.980-1.048)	0.447
	Traditionalist	0.081	1.084	(0.923-1.274)	0.326
Wealth Index	Poor	*Ref.*			
	Middle	0.052	1.054	(1.015-1.094)	0.006
	Rich	0.110	1.116	(1.066-1.170)	<0.001
Health Decision Maker	Wife	*Ref.*			
	Husband	-0.024	0.976	(0.949-1.004)	0.097
	Both	-0.001	0.999	(0.975-1.023)	0.918
	Others	0.075	1.078	(0.961-1.210)	0.202
Husband Edu. Level	No Edu.	*Ref.*			
	Primary	0.040	1.041	(0.995-1.088)	0.082
	Secondary	0.075	1.077	(1.038-1.119)	<0.001
	Higher	0.107	1.113	(1.067-1.162)	<0.001
Birth Order	1st	*Ref.*			
	2nd	-0.026	0.974	(0.927-1.023)	0.297
	3rd	-0.055	0.947	(0.897-0.999)	0.044
	4th	-0.068	0.935	(0.883-0.989)	0.018
	5th and above	-0.063	0.939	(0.878-1.004)	0.067
Place of Residence	Urban	*Ref.*			
	Rural	-0.049	0.952	(0.918-0.987)	0.007
Media Exposure	No	*Ref.*			
	Yes	0.052	1.054	(1.022-1.087)	<0.001
Ever Terminated Pregnancy	No	*Ref.*			
	Yes	0.035	1.036	(1.009-1.063)	0.008
Distance to Health Facility	Big problem	*Ref.*			
	Not a problem	0.039	1.040	(1.014-1.067)	0.003
Insurance Coverage	No	*Ref.*			
	Yes	0.077	1.080	(1.026-1.136)	0.003
Working Status	No	*Ref.*			
	Yes	0.034	1.035	(1.011-1.059)	0.004
Birth Type	Single	*Ref.*			
	Multiple	0.050	1.051	(0.984-1.122)	0.139

Table 6: Zero-inflation component results of ZINB – Factors associated with ANC non-visit

Predictors	Categories	Coeff.	AOR	95CI	P-value
Intercept	(Intercept	0.396	1.485	(0.958-2.301)	0.077
Maternal Age	15 – 24	*Ref.*			
	25 – 39	-0.153	0.858	(0.738-0.997)	0.046
	40 – 49	-0.153	0.858	(0.704-1.046)	0.129
Parity	1	*Ref.*			
	2 – 4	0.342	1.408	(1.103-1.798)	0.006
	5+	0.479	1.614	(1.181-2.208)	0.003
Region	North West	*Ref.*			
	North East	-1.005	0.366	(0.319-0.421)	<0.001
	North Central	0.811	2.250	(1.929-2.625)	<0.001
	South East	-0.218	0.804	(0.585-1.106)	0.180
	South-South	1.178	3.249	(2.453-4.304)	<0.001
	South West	0.535	1.708	(1.295-2.253)	<0.001
Maternal Edu. Level	No Edu.	*Ref.*			
	Primary	-0.635	0.530	(0.443-0.635)	<0.001
	Secondary	-1.059	0.347	(0.287-0.419)	<0.001
	Higher	-1.957	0.141	(0.087-0.228)	<0.001
Religion	Christian	*Ref.*			
	Islam	-0.109	0.897	(0.747-1.078)	0.246
	Traditionalist	-0.201	0.818	(0.457-1.465)	0.500
Wealth Index	Poor	*Ref.*			
	Middle	-0.313	0.731	(0.631-0.847)	<0.001
	Rich	-1.036	0.355	(0.277-0.455)	<0.001
Health Decision Maker	Wife	*Ref.*			
	Husband	0.260	1.297	(1.151-1.461)	<0.001
	Both	-0.254	0.776	(0.674-0.893)	<0.001
	Others	0.426	1.531	(0.974-2.405)	0.065
Husband Edu. Level	No Edu.	*Ref.*			
	Primary	-0.805	0.447	(0.373-0.536)	<0.001
	Secondary	-0.998	0.368	(0.315-0.431)	<0.001
	Higher	-1.585	0.205	(0.155-0.271)	<0.001
Birth Order	1^{st}	*Ref.*			
	2^{nd}	0.071	1.074	(0.834-1.383)	0.580
	3^{rd}	0.065	1.067	(0.809-1.406)	0.647
	4^{th}	0.175	1.191	(0.890-1.595)	0.239
	5^{th} and above	-0.060	0.941	(0.680-1.303)	0.716
Place of Residence	Urban	*Ref.*			
	Rural	0.196	1.217	(1.054-1.404)	0.007
Media Exposure	No	*Ref.*			
	Yes	-0.110	0.896	(0.738-1.088)	0.269
Ever Terminated Pregnancy	No	*Ref.*			
	Yes	-0.509	0.601	(0.513-0.704)	<0.001
Distance to Health Facility	Big problem	*Ref.*			
	Not a problem	-0.271	0.763	(0.687-0.847)	<0.001
Insurance Coverage	No	*Ref.*			
	Yes	-0.559	0.572	(0.238-1.374)	0.211
Working Status	No	*Ref.*			
	Yes	-0.445	0.641	(0.576-0.713)	<0.001
Birth Type	Single	*Ref.*			
	Multiple	-0.347	0.707	(0.488-1.023)	0.066

The zero-inflation component of the ZINB model results in Table 6 examined factors associated with non-attendance of antenatal care (ANC) visits. Maternal age was protective against ANC non-visit among women aged 25–39 (AOR = 0.858, $p = 0.046$) compared to those aged 15–24. Higher parity increased the likelihood of ANC non-attendance, with women having 2–4 children (AOR = 1.408, $p = 0.006$) and ≥5 children (AOR = 1.614, $p = 0.003$) being more likely to skip ANC visits. Regional disparities were evident: women in North East (AOR = 0.366, $p < 0.001$) were less likely to miss ANC visits, whereas women in North Central (AOR = 2.250, $p < 0.001$), South-South (AOR = 3.249, $p < 0.001$), and South West (AOR = 1.708, $p < 0.001$) had higher odds of non-attendance compared to those in North West.

Maternal education significantly decreased the likelihood of ANC non-visit, particularly for primary (AOR = 0.530, $p < 0.001$), secondary (AOR = 0.347, $p < 0.001$), and higher education (AOR = 0.141, $p < 0.001$). Husband's education also had a protective effect, with primary (AOR = 0.447, $p < 0.001$), secondary (AOR = 0.368, $p < 0.001$), and higher education (AOR = 0.205, $p < 0.001$) associated with lower odds of ANC non-attendance. Wealth status reduced the likelihood of non-visit, as women from middle- (AOR = 0.731, $p < 0.001$) and rich-class households (AOR = 0.355, $p < 0.001$) were less likely to skip ANC. Health decision-making by the husband increased the odds of ANC non-visit (AOR = 1.297, $p < 0.001$), while joint decision-making reduced this odds (AOR = 0.776, $p < 0.001$).

Other significant factors reducing ANC non-attendance included rural residence, which increased the odds (AOR = 1.217, p = 0.007), history of terminated pregnancy (AOR = 0.601, p < 0.001), shorter distance to health facilities (AOR = 0.763, p < 0.001) and employment (AOR = 0.641, p < 0.001). These findings underscore the importance of maternal and husband's education, parity, socioeconomic status, regional context, and access-related factors in determining ANC non-utilization.

The diagnostic tests in Table 7 indicated that the zero-inflated negative binomial (ZINB) model provides an adequate fit to the data. Residuals were approximately uniform (KS test, p = 0.103), and there was no evidence of over- or under-dispersion (dispersion = 1.037, p = 0.08). The zero-inflation test (ratio = 1.003, p = 0.752) suggested that the model appropriately accounted for excess zeros, and the frequency of outliers (0.0075, p = 0.583) was within the expected range. These results collectively indicate that the model assumptions are reasonable and the ZINB model is suitable for the data.

Table 7: ZINB model diagnostic result

Diagnostic Test	Statistic	P-value
Kolmogorov-Smirnov (Uniformity)	0.0108	0.1030
Dispersion	1.037	0.0800
Zero inflation	1.003	0.7520
Outlier Frequency	0.0075	0.5830

The Q-Q plot in Fig. 2 indicates that the random effects do not deviate significantly from a normal distribution. The points adhere closely to the theoretical line, satisfying this model assumption.

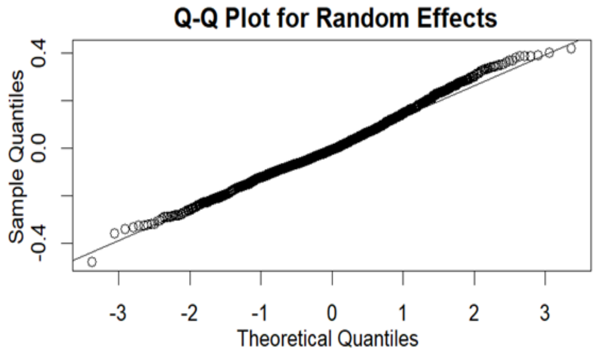

Figure 2: Q-Q plot for random effects

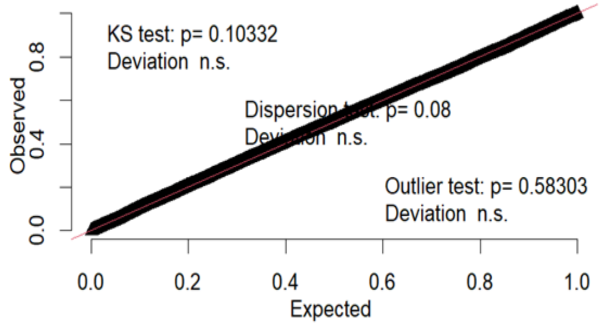

Figure 3: Q-Q plot of residuals

The Q-Q plot of residuals in Fig. 3 indicates that the residuals do not significantly deviate from a normal distribution. This is supported by non-significant results from a Kolmogorov-Smirnov test (p = 0.103), a dispersion test (p = 0.08), and an outlier test (p = 0.583). The assumption of normality of residuals is therefore satisfied.

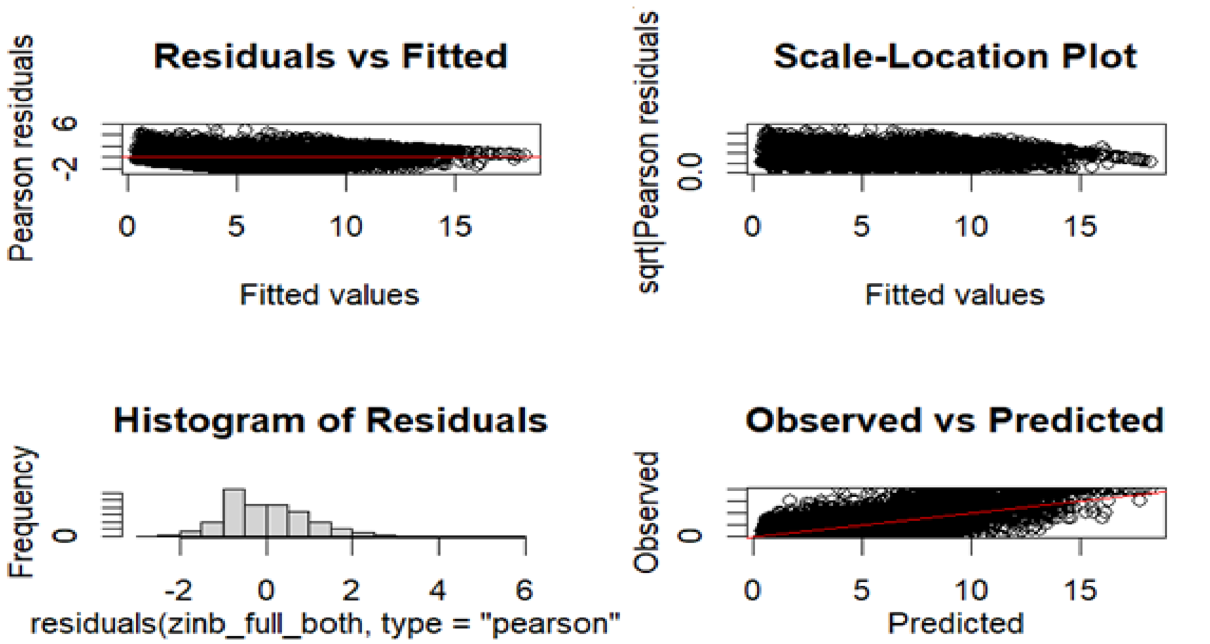

Figure 4: Residual diagnostic plots for the ZINB model

The residuals diagnostic plots in Fig. 4 indicated that the ZINB model meets key assumptions. The Residuals vs. Fitted and Scale-Location plots show no pattern, confirming linearity and homoscedasticity. The Observed vs. Predicted values demonstrate strong predictive accuracy. The distribution of residuals is approximately normal, supporting the model's validity.

CONCLUSION

This study demonstrates that antenatal care (ANC) utilization in Nigeria is influenced by a complex interplay of demographic, socioeconomic, educational and access-related factors. Maternal and husband education, household wealth, maternal age, parity, place of residence, regional context, media exposure, employment, insurance coverage, and proximity to health facilities were all significant determinants of both the frequency of ANC visits and the likelihood of non-attendance. The findings underscored substantial regional disparities and highlight the vulnerability of high-parity and rural women to inadequate ANC.

Methodologically, the use of a multilevel Zero-Inflated Negative Binomial (ZINB) model effectively accounted for overdispersion, excess zeros, and clustering in the data, providing robust and reliable estimates. This approach confirms the importance of appropriately modeling hierarchical count data in maternal health research, as demonstrated in comparable studies across Bangladesh and Ethiopia.

Policy and programmatic interventions should therefore prioritize educational empowerment, socioeconomic support, improved health facility accessibility and targeted regional strategies to increase ANC utilization. Future research should explore the integration of quality-of-care metrics and longitudinal analyses to better understand barriers and facilitators of maternal health service uptake.

This study leveraged on a nationally representative dataset and employed a multilevel Zero-Inflated Negative Binomial (ZINB) model, effectively accounting for overdispersion, excess zeros, and the hierarchical structure of the data. By analyzing both the count and zero-inflation components, the study provided a comprehensive assessment of factors influencing both the frequency of antenatal care (ANC) visits and the likelihood of non-attendance. The inclusion of demographic, socioeconomic, regional, and access-related variables allowed for nuanced insights into disparities in ANC utilization across Nigeria.

The cross-sectional design limits causal inferences and reliance on self-reported data may introduce recall or social desirability bias. Important factors such as quality of care, cultural practices and health facility characteristics were not captured, which may influence ANC utilization. Additionally, while model diagnostics indicate good fit, the ZINB model assumes conditional independence between clusters, which may not fully account for all hierarchical dependencies in the data.

Conflict of interest: The authors declare no conflict of interest.

References

Ahmed, T. and Mallick, S. K. (2019). Analyzing antenatal care visit data in Bangladesh using zero-modified count regression models. *Journal of Health Research*, 33(5), 412–425.

Beckman, A., Chaix, B., Hjerpe, P., Johnell, K., Larsen, K., Merlo, J., Ohlsson, H. and Råstam, L. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health*, 60(4), 290–297. https://doi.org/10.1136/jech.2004.029454

Bekalo, D. B. and Kebede, D. T. (2021). Zero-inflated models for count data: An application to number of antenatal care service visits. *Annals of Data Science*, 8(4), 683–708.

Bhowmik, J., Biswas, R. K. and Woldu, M. (2020). Determinants of antenatal care visits among pregnant women in Bangladesh: An advanced count regression modeling. *PLOS ONE*, 15(11), e0243165.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. and White, J. S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M. and Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge University Press. https://doi.org/10.1017/CBO9780511814365

Consul, P. C. and Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics*, 15(4), 791–799. https://doi.org/10.1080/00401706.1973.10489112

Gebeyehu, N. A., Arage, G. and Degu, A. (2022). Antenatal care visits and associated factors in Sub-Saharan Africa: A multilevel negative binomial regression analysis. *BMC Pregnancy and Childbirth*, 22(1), 1–12.

Hilbe, J. M. (2011). *Negative Binomial Regression* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511973420

Hox, J. J., Moerbeek, M. and Van de Schoot, R. (2017). *Multilevel Analysis: Techniques and Applications* (3rd ed.). Routledge. https://doi.org/10.4324/9781315650982

Kreft, I. G. and de Leeuw, J. (1998). *Introducing Multilevel Modeling*. Sage Publications. https://doi.org/10.4135/9781849209366

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1–14. https://doi.org/10.2307/1269547

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3), 341–365. https://doi.org/10.1016/0304-4076(86)90002-3

Negash, W., Tefera, L. and Abate, M. (2025). Antenatal care visits and associated factors in Senegal: A multilevel Poisson regression analysis of the 2023 DHS. *Maternal and Child Health Journal*, 29(1), 150–162.

Payne, E. H., Gebregziabher, M., Hardin, J. W., Ramakrishnan, V. and Egede, L. E. (2018). An empirical approach to determine a threshold for assessing overdispersion in Poisson and negative binomial models for count data. *Communications in Statistics - Simulation and Computation*, 47(6), 1722–1738. https://doi.org/10.1080/03610918.2017.1323223

Sekata, F. (2015). Antenatal care visits and associated factors in rural Ethiopia: Application of zero-inflated and hurdle regression models. Master's thesis, University of Oslo.

Suleman, M., Yaya, S. and Zegeye, B. (2023). Factors associated with the desired number of antenatal care service visits in Ethiopia: A multilevel count regression analysis. *BMJ Open*, 13(2), e067891.

Woldeyohannes, D., Asnake, M. and Teklu, A. (2023). Determinants of antenatal care utilization in Ethiopia: A count regression approach. *Ethiopian Journal of Health Sciences*, 33(1), 1–12.

World Health Organization (2016). WHO Recommendations on Antenatal Care for a Positive Pregnancy Experience. WHO Press.

Yadav, A. K., Nag, S., Jena, P. K. and Paltasingh, T. (2021). Micro-level determinants of antenatal care utilization in India: A count data modeling approach. *Children and Youth Services Review*, 120, 105780.

Yusuf, O. B. and Ugalahi, T. O. (2015). Comparison of Poisson, negative binomial, and generalized Poisson regression models in predicting antenatal care visits in Nigeria. *Journal of Biometrics & Biostatistics*, 6(6), 1–7.

Yusuf, O. B., Afolabi, R. F. and Agbaje, A. S. (2018). Comparing the performance of zero-inflated and hurdle models for antenatal care utilization in Nigeria. *BMC Medical Research Methodology*, 18(1), 1–12.

Zemene, M. and Birhanu, A. (2022). Factors affecting antenatal care utilization in Ethiopia: A count regression model analysis of the 2019 EMDHS. *Journal of Pregnancy*, 1–10.