© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712685

Model Comparison of Logistic, Poisson, And Negative Binomial Regression Approaches for Determining Predictors of Mother-To-Child Transmission of Syphilis in Ebonyi North Senatorial District.

NWUZOR OZOEMENA[1], OKORO CHIEMEKA NWANKWOR[2], S. C NWASUKA[3]
[1]*Department of Industrial Mathematics and Applied Statistics, Ebonyi State University Abakaliki, Nigeria.*
[2]*Department of Industrial Mathematics and Applied Statistics, Ebonyi State University Abakaliki, Nigeria.*
[3]*Department of Mathematics and Computer Science, Clifford University, Owerrinta.*

***Abstract:** This study examined factors influencing mother-to-child transmission (MTCT) of syphilis in Ebonyi North Senatorial District and compared the predictive performance of logistic, Poisson, and negative binomial regression models. Data from four healthcare facilities (2010–2024) were analyzed. Key service indicators, including drug availability, screening, and counseling uptake, were assessed. Model comparison results showed that the logistic regression model provided the strongest predictive power, with excellent fit indices (McFadden $R^2 = 0.92$; AUC = 0.998). Even after variable reduction using PCA, the logistic model maintained high performance (McFadden $R^2 = 0.63$; AUC = 0.964). In contrast, the Poisson and negative binomial models identified similar significant predictors—age and syphilis knowledge—but the negative binomial model outperformed Poisson due to better handling of overdispersion. Overall, logistic regression demonstrated superior accuracy and discriminatory ability compared to the count-based models. The findings highlight logistic regression as the most reliable model for identifying MTCT risk factors while emphasizing the need to strengthen screening, counseling, and treatment services to reduce congenital syphilis.*

***Keywords:** Syphilis, Transmission, Influencing, Logistic, Poison and Negative Binomial Regression Models.*

I. INTRODUCTION

Syphilis, a sexually transmitted infection caused by "Treponema pallidum", remains a major global public health challenge. The infection can be acquired through sexual contact, blood transfusion, or vertically from an infected mother to her fetus during pregnancy (WHO, 2024). Mother-to-child transmission (MTCT), known as congenital syphilis, represents one of the most severe pathways of transmission (Araujo et al., 2013). Congenital syphilis can result in devastating outcomes such as stillbirth, neonatal mortality, and long-term neurological or developmental impairments among affected infants (Garnett et al., 2019).

Recent scholarly attention has focused on understanding MTCT of syphilis through the application of mathematical and statistical modeling techniques. Diagnosis typically involves serological tests detecting antibodies to "T. pallidum", while treatment relies primarily on penicillin administration. Logistic regression is frequently employed to assess binary outcomes, such as the presence or absence of congenital infection, whereas Poisson regression is commonly used for analyzing count data, including reported cases within populations (Smith and Johnson, 2015). The negative binomial regression model—an extension of the Poisson model—accounts for overdispersion, a frequent occurrence in infectious disease data. These advanced statistical approaches enable researchers to identify key risk factors for congenital syphilis, including maternal age, infection stage, and other clinical or demographic variables.

Globally, syphilis persists as a significant burden, with an estimated 930,000 pregnant women affected annually, contributing to adverse maternal and neonatal outcomes (Newman et al., 2013). If untreated, MTCT can lead to severe complications, including stillbirth, neonatal death, and long-term disabilities (Dorling et al., 2019). In Ebonyi State, Nigeria, approximately 58,320 individuals are estimated to be living with syphilis out of a total population of 3.5 million, with a reported prevalence of 1.8% among pregnant women (Federal Ministry of Health, 2018). Despite this burden, limited research in the region has employed logistic, Poisson, or negative binomial models to analyze MTCT patterns. Most existing studies focus on individual risk factors rather than leveraging advanced statistical tools to understand transmission dynamics comprehensively.

This study seeks to address this gap by identifying the determinants of MTCT of syphilis in Ebonyi State and developing predictive models capable of estimating transmission risk based on these factors. The aim of the research is to investigate the factors influencing mother-to-child transmission of syphilis using logistic, Poisson, and negative binomial regression models.

II. METHODOLOGY/ RESEARCH ELABORATIONS

2.1: Study Area

The study was conducted in four general hospitals within Ebonyi North Senatorial District of Ebonyi State, Nigeria. Ebonyi North Senatorial District encompasses the northern part of Ebonyi State, Nigeria. It includes the Local Government Areas (LGAs) of Abakaliki, Izzi, Ebonyi, Ohaukwu, and part of Ezza North. The district is primarily located in the Abakaliki division of the state. Ebonyi North Senatorial District is one of the three senatorial districts in Ebonyi State, the others being Ebonyi Central and Ebonyi South. The district is predominantly Igbo-speaking and shares borders with Enugu State to the north and northwest.

2.2 Data Collection Method

This study is a cross-sectional quantitative survey carried out using interviewer administered structured questionnaires and secondary data method. Secondary data were collected using records of tested pregnant women and number of observed positives of syphilis patients which also captured variables in the conceptual framework, as this allowed the study to appropriately measure those factors that were important in answering the research objectives. The data captured data about demographic characteristics and background Information such as: age, marital status educational level, diagnosis method, number of survived children after delivery, number of children with reactive status, mode of transportation and time taken to get to the facility

2.3 Methodology

The first step in conducting this research involved defining the target population, which consisted of the four hospitals previously identified. The second step was obtaining access to existing data sources and reports on syphilis testing among pregnant women, which provided the samples for analysis. The third step involved categorizing the population into relevant subgroups before carrying out the analysis and interpreting the results.

Three statistical models were employed in this study for data analysis: the Negative Binomial Regression Model (NBRM), the Poisson Regression Model (PRM), and the Logistic Regression Model (LRM). The binomial distribution is defined as a discrete probability distribution that describes the number of successes in a fixed number of independent and identically distributed trials.

The probability Mass Function of Negative Binomial Model is given by:

$$pr(X = x) = \binom{k+x-1}{x-1} p^k q^x \quad 0 \leq x < \infty \qquad 1$$

Where p is the probability of success and q =1-p is the probability of failures
The mean is given by E(X) $= \mu = \alpha + \beta X$, the variance is Var(X) = $\mu + (\frac{1}{\phi}) * \mu^2$ and the moment generating function: $M_x(t) = \frac{(1-\phi *)^{-\alpha}}{(1-\varphi *(t+\beta)^{-\alpha}}$
Poisson distribution is a probability distribution that represents the number the number of events occurring in a fixed interval of time or space. It is often used to model rare events that occur randomly such as the number of phone calls received at a call center in a given hour or the number of cars passing through a toll booth in a given minute.
The probability mass function of a poison regression of a variable X is given by :

$$\Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} x = 0,1, \ldots\ldots \infty \qquad 2$$

The mean and the variance of the distribution are both λ and the moment generating function: $M_x(t) = e^{\lambda(e^t - 1)}$. The fundamental Poisson regression model for an observation i is written as:

$$\Pr(Y_i = y_i | u_i, t_i) = \frac{e^{-\mu i t_i}(\mu_{it_i})^{y_i}}{y_{i!}} \quad 3$$

Logistic regression in the other hand is also known as the logistic model or logit model is used in this research to analyze the relationship between multiple independent variables and a categorical dependent variable, and estimates the probability of occurrence of an event by fitting data to logistic curve.

Basically, there are two models of logistic regression which include: Binary logistic regression and multinomial logistic regression. Binary logistic regression is basically used when the dependent variable is dichotomous and the independent variables are either continuous or categorical. But when the dependent variable is not dichotomous and is composed of more than two categories, then a multinomial logistic regression can be employed (Kleinbaum & Klien, 2010). Logistic regression can be used for various growth models, and is used in a certain type of regression regarded as logistic regression model (Lawal, 2003). And the probability density function is given by:

$$F(y) = \frac{e^{-(x-\mu)}}{[\sigma \times 1 + e^{-(x-\mu)/\sigma)^2}}, \text{ where } x \in \mathbb{R} \quad 4$$

Odds are the ratio of the probability of an event occurring to the probability of it not occurring. If the probability of an event occurring is p, the probability of the event not occurring is (q). Then the corresponding odds are a value given by:

$$\text{Odds of event} = \frac{p}{q} \quad 5$$

III. RESULTS AND DISCUSSION

Table 1 – Analysis of deviance results of the logistic regression

	Df	Deviance	Resid.DF	Resid.Dev	Pr(>Chi)
NULL			235	326.080	
x1	3	10.899	232	315.180	0.012
x2	1	0.781	231	314.400	0.377
x3	2	2.550	229	311.850	0.279
x4	2	25.668	227	286.180	0.000
x5	2	11.302	225	274.880	0.004
x6	2	59.246	223	215.630	0.000
x7	2	11.205	221	204.430	0.004
x8	2	22.727	219	181.700	0.000
x9	2	2.014	217	179.690	0.365
x10	2	5.108	215	174.580	0.078
x11	2	23.132	213	151.450	0.000
x12	2	4.800	211	146.650	0.091
x13	2	79.003	209	67.650	0.000
x14	2	11.422	207	56.220	0.003
x15	2	3.604	205	52.620	0.165
x16	2	1.588	203	51.030	0.452

x17	2	1.947	201	49.090	0.378
x18	2	0.000	199	49.090	1.000
x19	2	0.000	197	49.090	1.000
x20	2	0.000	195	49.090	1.000
x21	2	4.287	193	44.800	0.117
x22	1	5.062	192	39.740	0.024
x23	2	0.000	190	39.740	1.000
x24	2	4.499	188	35.240	0.105
x25	1	0.000	187	35.240	1.000
x26	1	1.185	186	34.050	0.276
x27	2	0.000	184	34.050	1.000
x28	2	0.000	182	34.050	1.000
x29	2	1.726	180	32.330	0.422
x30	1	0.505	179	31.820	0.477
x31	2	0.000	177	31.820	1.000
x32	2	0.000	175	31.820	1.000
x33	2	4.499	173	27.320	0.105
x34	2	0.000	171	27.320	1.000
x35	2	0.000	169	27.320	1.000

This table shows that the significant predictors include maternal age, exposure to syphilis patients, blood transfusion history, pregnancy status, presence of syphilis symptoms, other STIs, screening, counselling, infant treatment, and use of protection while the non-significant predictors include marital status (x2), knowledge (x3), breastfeeding transmission (x16), spousal permission (x24), stigma (x27), and others did not show significant influence in this model.

Table 2 – Goodness of tests results:

	Chi-square	p-value
Wald	4.77	1
Hosmer-Lemeshow	2.66E-08	1

This table shows that the logistic regression model fits the data extremely well, with the Hosmer–Lemeshow test confirming excellent calibration ($p = 1.000$). Although the Wald test did not show significance, the strong fit suggests the predictors reliably explain MTCT of syphilis.

Table 3 – Psuedo R^2 values for PCA selected independent variables

	Value
Llh	-61.0680
LlhNull	-163.0399
G2	203.9437
McFadden	0.6254
r2ML(Cox & Snell)	0.5785
r2CU (Nagelkerke)	0.7726

Table 3 demonstrates that the logistic regression model fits the data very well (Nagelkerke $R^2 = 0.77$, McFadden $R^2 = 0.63$). This means the factors identified (through PCA ie principal component analysis) strongly influence MTCT of syphilis in Ebonyi North.

© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712685

Table 4– Goodness of tests results for PCA selected independent variables

	Chi-square	p-value
Wald	44.271	0.0008
Hosmer-Lemeshow	20.669	0.0080

The Wald test checks whether the selected independent variables (from PCA) significantly contribute to explaining mother-to-child transmission (MTCT) of syphilis.

Since p = 0.0008 < 0.05, this indicates the independent variables included in the logistic regression model are statistically significant predictors of MTCT. Also, Hosmer-Lemeshow test evaluates the goodness of fit of the logistic regression model. Ideally, a non-significant result (p > 0.05) means the model fits well. But here, p = 0.0080 < 0.05, meaning there is a poor fit between the predicted probabilities and the observed outcomes.

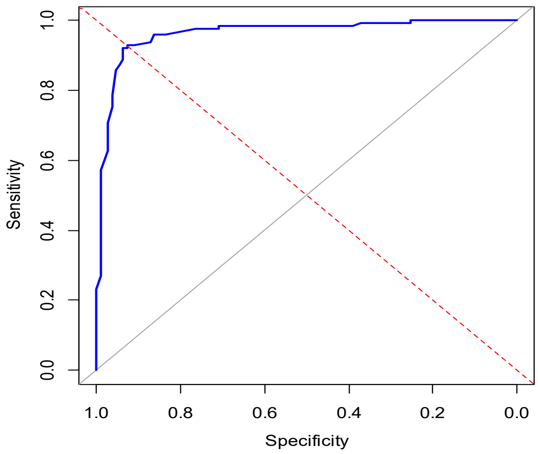

Fig. 1 - ROC Curve for Logistic Regression using PCA selected independent variables (AUC = 0.9637)

The blue line represents the ability of the logistic regression model to correctly classify whether mother-to-child transmission (MTCT) of syphilis occurred or not. It plots the True Positive Rate (Sensitivity) against the False Positive Rate (1 – Specificity) at different threshold values. The red diagonal line is the line of no discrimination meaning it has no predictive power. The blue ROC curve rises steeply towards the top-left corner, which indicates very high sensitivity and specificity. This means the model is excellently distinguishing between mothers who transmitted syphilis to their babies and those who did not.

Table 5 – Poisson regression results

	Estimate	Std. Error	z-value	2.50%	97.50%	IRR	Pr(>|z|)
(Intercept)	3.934	7.122	0.552	-10.0287	17.89214	51.118	0.581
X1	-0.150	0.017	-8.803	-0.18391	-0.11695	0.860	<2e-16
X2	-0.004	0.004	-1.269	-0.01141	0.002444	0.996	0.205
X3	1.102	0.072	15.349	0.961192	1.242641	3.010	<2e-16

This table shows X1(Age) significantly reduces MTCT risk (protective factor).X2(Marital status) shows that it has no significant effect while X3(Syphilis knowledge) significantly increases reported MTCT, possibly due to better awareness, higher testing uptake, or concentration of knowledge among high-risk women.

Table 6 – Negative Binomial results

	Estimate	Std. Error	z-value	2.50%	97.50%	IRR	Pr(>|z|)
(Intercept)	1.714	24.382	0.070	-46.215	49.616	5.553	0.944
X1	-0.147	0.057	-2.572	-0.254	-0.041	0.863	0.010
X2	-0.003	0.012	-0.286	-0.027	0.020	0.997	0.775
X3	1.122	0.251	4.475	0.628	1.622	3.072	0.000

This table shows that X1(Age) is a significant protective factor — younger mothers need more targeted interventions meaning that younger mothers are at higher risk of transmitting syphilis to their babies compared to older mothers that is for each one-unit increase in age, the incidence rate of MTCT decreases by about 13.7%((1 – 0.863). Also, X2(Marital status) is not a significant factor in MTCT risk because this suggests that whether a woman is single, married, divorced, or widowed doesn't meaningfully alter the transmission rate. And X3(Knowledge) alone is not sufficient — though women may know about syphilis, structural and behavioral barriers still allow MTCT to occur because Women with higher syphilis knowledge have a 3.07 times higher incidence rate of MTCT compared to those with less knowledge.

IV. CONCLUSION

This study shows that mother-to-child transmission of syphilis in Ebonyi State is largely driven by maternal characteristics (such as age, pregnancy status, and overall health), service-delivery challenges (including drug availability, follow-up counseling, and client satisfaction), and socio-cultural constraints (notably stigma and the need for spousal approval). Among the models applied, logistic regression using PCA-selected variables produced the most reliable predictions, while the negative binomial model reinforced the significance of maternal age and syphilis knowledge as key determinants. Although MTCT-prevention measures exist in the assessed health facilities, persistent gaps—particularly in drug supply, counseling quality, and follow-up support—continue to hinder effective prevention. Addressing these structural weaknesses is essential if the goal of eliminating congenital syphilis in Ebonyi State is to be achieved.

REFERENCES

[1] Araújo, C.L., Gonçalve, CV, Pires, R.D., and Araújo, T.M.(2013). Factors associated with syphilis in pregnant women: a case-control study.At the Hospital Maternidade Leonor Mendes de Barros, Sao Paulo, Brazil .*Rev Bras Ginecol Obstet*. (1):22-7.

[2] Bartlett, J. E., Kotlrik, J. W. & Higgigs, C. C (2001): Organizational Research: DeterminingAppropriate Sample size in survey Research. Learning and Performance Journal, 43 – 50.

[3] Bell, A. and Marshall, A.J. (1964). Serological diagnosis of syphilis. *British Journal of Venereal Diseases.40,*45-49.

[4] Blocker, M. E., Levine, W. C., and Louis, M. E. S. (2000). HIV prevalence in patients with syphilis, United States. *Sexually transmitted diseases*, *27*(1); 53-59.

[5] Brandt, A. M. (1978). Racism and research: the case of the Tuskegee Syphilis Study. *Hastings center report*, 21-29.

[6] Brown, A., & Lee, S. (2017). Understanding the Negative Binomial Regression Model in Marketing Research. *Journal of Marketing Research, 20*(4), 345-358.

[7] Brown, A., and Smith, J. (2020). Assessing the Impact of Independent Variables on Firm Performance using Logistic Regression Models. *Journal of Business Management, 35*(4), 312-326.

[8] Brown, R. J., Araujo-Vilar, D., Cheung, P. T., Dunger, D., Garg, A., Jack, M., ... and Yorifuji, T. (2016). The diagnosis and management of lipodystrophy syndromes: a multi-society practice guideline. *The Journal of Clinical Endocrinology and Metabolism*, *101*(12), 4500-4511

[9] Brown, R., Gagnon, R., and Delisle, M. F. (2019). No. 373-cervical insufficiency and cervical cerclage. *Journal of Obstetrics and Gynaecology Canada*, *41*(2), 233-247.

[10] Dorling, J., Abbott, J., Berrington, J., Bosiak, B., Bowler, U., Boyle, E., ...and Townend, J. (2019). Controlled trial of two incremental milk-feeding rates in preterm infants. *New England Journal of Medicine*, *381*(15), 1434-1443.

[11] Garnett, G. P., and Aral, S. O. (2018). The resurgence of syphilis: a literature review. *Sexually Transmitted Infections, 25*(2), 89-96.

[12] Lawal, H.B (2003): Categorical Data analysis with SAS and SPSS Applications. London: Lawrence Erlbaum Associates, publishers.

[13] Newman, L., Kamb, M., Hawkes, S., Gomez, G., Say, L., Seuc, A and Broutet, N.(2013) Global estimates of syphilis in pregnancy and associated adverse outcomes: analysis of multinational antenatal surveillance data. *Public Library of Science Medicine(PLoS). 10*(2):e1001396.

[14] Nguyen, V.T.T., Trang, H.T.Q, and Ishikawa, N. (2021).Feasibility Benefits and Cost-effectiveness of Adding Universal Hepatitis B and Syphilis Testing to Routine Antenatal Care Services in Thai Province, Vietnam. *International Journal of STD and Aids.32*(2);135-143.

[15] Nwuzor, O., Okoro, C.N., and Igwe , S. (2025). Investigating the factors influencing mother-to child transmission of syphilis in Ebonyi North Senatorial District using logistic, Poisson, and negative binomial regression models.Iconic Research and Engineering Journals, 9(5), ISSN 2456-8880.

[16] Smith, J., and Johnson, R. (2015). A comparison of logistic and poison regression models for predicting disease outbreaks. *Journal of Epidemiology, 10*(2), 123-135.

[17] World Health Organization (2024). *Implementing the global health sector strategies on HIV, viral hepatitis and sexually transmitted infections, 2022–2030: report on progress and gaps 2024*. World Health Organization.

[18] Zhou, P., Ye, M., Tucker, J. D., Zhang, L., and Radolf, J. D. (2023). Syphilis infection: clinical, epidemiology, basic science, and behavioral research. *Frontiers in Immunology*, *14*, 1182069.

[19] Xiao, Y., Li, S. L., Lin, H. L., Lin, Z. F., Zhu, X. Z., Fan, J. Y., ... and Yang, T. C. (2016). Factors associated with syphilis infection: a comprehensive analysis based on a case-control study. *Epidemiology and Infection*, *144*(6), 1165-1174.

[20] Xu, J. J., Reilly, K. H., Lu, C. M., Ma, N., Zhang, M., Chu, Z. X., ... and Shang, H. (2011). A cross-sectional study of HIV and syphilis infections among male students who have sex with men (MSM) in northeast China: implications for implementing HIV screening and intervention programs. *BMC Public Health*, *11*, 1-8.

[21] Yang, S., and Berdine, G. (2015). The negative binomial regression. *The Southwest Respiratory and Critical Care Chronicles*, *3*(10), 50-54.

[22] Yin, Y. P., Ngige, E., Anyaike, C., Ijaodola, G., Oyelade, T. A., Vaz, R.