Economics and Statistics Research Journal (ESRJ)
Volume.16, Number 7; July-2025;
ISSN: 2836-8398 | Impact Factor: 9.97
https://zapjournals.com/Journals/index.php/esrj
Published By: Zendo Academic Publishing

MODELLING THE DETERMINANTS AND FORMS OF DOMESTIC VIOLENCE AGAINST PREGNANT WOMEN: A MULTINOMIAL LOGISTIC REGRESSION APPROACH TO CATEGORICAL DATA ANALYSIS

[1*]Gongden M.S, [2]Nweze N.O, [2]Bilkisu M. and [2]Abdulahi Abubakar
Email: gongdenmicah@gmail.com
Phone: +2348037050376

Article Info
Keywords: Multinomial Logistic Regression, Domestic Violence During Pregnancy, Public Health Modeling, Socio-behavioral Risk Factors, Maternal Health Interventions
DOI
10.5281/zenodo.16318837

Abstract
This study applies a Multinomial Logistic Regression (MLR) framework to model unordered categorical outcomes in public health research, to examine forms of domestic violence experienced during pregnancy. MLR was selected for its capacity to estimate the probability of multiple, non-ordinal outcome categories such as emotional, physical, and sexual violence relative to a set of predictor variables (Agresti, 2018; Hosmer, Lemeshow, & Sturdivant, 2013). Using cross-sectional data from 499 pregnant women attending antenatal care, the model assessed the influence of maternal education, employment status, partner substance use, and mental health history on the likelihood of experiencing each form of violence. MLR results revealed that emotional violence was the most prevalent, followed by physical and sexual violence. The study demonstrates the effectiveness of MLR in disentangling complex, non-binary health outcomes and highlights the pressing need for integrated psycho-social screening and context-specific intervention strategies within maternal health programs (World Health Organization [WHO], 2013; Campbell, 2002). It also supports policy development aimed at addressing the socio-behavioral risk factors that heighten women's vulnerability to domestic violence during pregnancy (Barnett et al., 2011; Bandura, 1977).

1.0 Introduction

Understanding the determinants of public health outcomes involving non-ordinal categorical responses requires robust statistical frameworks that go beyond conventional binary classification. Multinomial Logistic Regression (MLR) is a powerful modeling technique suited for such analyses, as it allows for the estimation of the

[1] PhD Research Student, Department of Statistics and Data Analytics, Nasarawa State University Keffi, Nigeria.
[2] Lecturer, Department of Statistics and Data Analytics, Nasarawa State University Keffi, Nigeria

probabilities of multiple, mutually exclusive outcome categories based on a set of explanatory variables. Unlike binary logistic regression, MLR does not assume any inherent order among outcome categories and enables researchers to simultaneously evaluate the influence of predictor variables on each outcome relative to a chosen reference group. This methodological strength makes MLR particularly appropriate for examining complex social phenomena such as Domestic Violence (DV), which may manifest in various forms—emotional, physical, or sexual—that are conceptually distinct but equally consequential. Traditional analyses of DV often rely on binary outcomes (e.g., violence present vs. not present), which obscures the heterogeneity of experiences and prevents precise risk profiling for intervention purposes.

2.0 Problem Statement

Domestic violence remains a pervasive global concern, with the World Health Organization (WHO) 2002, estimating that one in three women globally has experienced some form of abuse. The vulnerability is even more acute during pregnancy, where physiological, psychological, and relational dynamics heighten risk. In Nigeria, the issue is compounded by cultural norms, social stigma, and limited institutional support, contributing to the underreporting and normalization of DV in maternal contexts. This study leverages MLR to model the probability of experiencing specific forms of domestic violence among pregnant women attending antenatal clinics at Plateau State Specialist Hospital, Jos. The outcome variable comprises three nominal categories: emotional, physical, and sexual violence. Predictor variables include sociodemographic and psychosocial factors such as education level, employment status, partner's substance use, and mental health history.

In the context of modeling domestic violence among pregnant women, probability theory underpins the use of multinomial logistic regression to analyze categorical dependent variables with multiple possible outcomes. This regression technique models the probability that a given individual falls into one of several discrete categories—such as experiencing emotional, physical, sexual violence, or no violence—conditional on predictor variables (Hosmer, Lemeshow, & Sturdivant, 2013).

The dependent variable in Multinomial Logistic Regression follows a multinomial probability distribution, which generalizes the binomial distribution to cases where each trial can result in more than two outcomes (Agresti, 2018). The model estimates the probability of each outcome category as a function of explanatory variables, typically using the maximum likelihood estimation method (MLE) to find the parameter values that maximize the probability of observing the sample data (Hosmer et al., 2013). For example,

$P(Y = j) = \beta_0 + \beta_1 X_1 + \beta_2 \beta X_2 + ... + \beta_K X_K$

Where P (Y = j) is the probability of experiencing type j of domestic violence, P(Y=0) ,P(Y = 1) and P(Y=2) is the probability of no domestic violence (reference category), and $X_1, X_2, X_3 \dots, X_K$ is the probability of no domestic violence (reference category), and represent the independent variables age, education, or partner's behavior

probability theory allows researchers to estimate the conditional probability by linking predictors to probabilities through a logit (log-odds) transformation, the model provides interpretable coefficients that describe how changes in predictors influence the likelihood of each violence type relative to a baseline category (Menard, 2002). Probability theory is the mathematical foundation for modeling uncertainty and random phenomena. It provides the framework to quantify how likely events are to occur and to infer relationships between variables when outcomes are not deterministic.

In the context of domestic violence research among pregnant women, probability theory underpins statistical models like multinomial logistic regression, which estimates the probabilities of different categories of a categorical outcome—in this case, various forms of domestic violence (e.g., emotional, physical, sexual).

The objectives of the study are to estimate the prevalence of each form of violence during pregnancy, and identify significant predictors associated with each category using a model-driven, statistically sound approach. By aligning methodological rigor with a critical health and social problem, this research contributes both substantively to the understanding of DV in pregnancy and methodologically to the application of categorical data analysis in public health. Most empirical studies on DV adopt a binary perspective measuring violence as either present or absent without distinguishing between different types. This approach fails to capture the nuanced ways in which emotional, physical, and sexual abuse manifest and intersect Ogbo et al. (2018). Given that these forms of violence are non-ordinal and often occur independently, there is a need for analytic techniques that reflect this complexity.

This study adopted a descriptive cross-sectional design, consistent with the requirements for applying Multinomial Logistic Regression (MLR) to categorical outcome data. The design facilitated the collection of individual-level data at a single time point, allowing for the identification of predictors associated with different forms of domestic violence without assuming longitudinal relationships. Data were collected from Plateau State Specialist Hospital, Jos, a major tertiary healthcare institution in North Central Nigeria. The hospital provides comprehensive maternal and child health services to both urban and rural populations, making it a suitable setting for studying pregnant women from diverse sociodemographic backgrounds.

Methodology

The sample size was calculated using Cochran's formula for cross-sectional studies:

$$n = \frac{z^2 pq}{d^2}$$

Where

n = Minimum sample size

Z = Area under normal curve corresponding to 95% confidence interval.

= 1.96

P = Expected prevalence

= 0.5

q = 1 – P

d = Desired precision

= 0.05 (for 95% confidence interval)

Therefore, n = **384**

With a 0.05 margin of error. After adjusting for potential non-response, the final sample size was determined to be approximately 400. The target population comprised pregnant women attending antenatal clinics at the study site. Using systematic random sampling, a total of 400 respondents were selected over a three-month period. Eligibility criteria included being pregnant, aged 18 years or older, and providing informed consent. The sample size was adequate for MLR analysis, which requires sufficient observations across multiple outcome categories to ensure stable and interpretable estimates**.**

Study Design

A descriptive cross-sectional design was employed. The study was conducted at Plateau State Specialist Hospital, Jos, a tertiary healthcare facility offering maternal services to urban and rural populations in North-Central Nigeria.

Sample Size and Sampling Technique

The sample size was determined using Cochran's formula for cross-sectional studies (Iliyasu et al., 2013), where p = 0.37 (based on Iliyasu et al., 2013). After adjusting for non-response, the final sample was 499 pregnant women selected through systematic random sampling.

Data Collection Instrument

A structured, interviewer-administered questionnaire was developed based on standardized tools, including the World Health Organization (WHO 2023) It included sections on experiences of emotional, physical, and sexual violence; sociodemographic characteristics; and partner behavior and mental health history (Rabin et al., 2009). The instrument captured data on:

- Self-reported experiences of emotional, physical, and sexual violence
- Sociodemographic characteristics (age, education, employment, household income)
- Partner characteristics (education level, substance use)
- Maternal mental health history

The dependent variable was the type of domestic violence experienced, operationalized as a nominal categorical variable suitable for MLR analysis. Variables and Operationalization. Dependent Variable: *Form of domestic violence*: A categorical variable with three unordered levels—emotional, physical, and sexual—based on the primary type of violence reported by each respondent.

Independent Variables: Maternal age (years), Maternal education level (none, primary, secondary, tertiary), Employment status (employed, unemployed), Partner's education level, Partner's substance use (yes/no), Household income (monthly income categories), History of mental health problems (yes/no), Number of children (parity)

These variables were selected based on theoretical relevance and prior empirical evidence on risk factors for intimate partner violence. They were all coded appropriately for inclusion in the MLR model.

Variables and Operationalization

Dependent Variable: Form of domestic violence (emotional, physical, sexual). **Independent Variables**: Maternal age, education level, employment status, partner's education, partner's substance use, household income, mental health history, and parity (Barnett et al., 2011)

Statistical Analysis.

We consider the linear form of the regression specified as

We consider the linear form of the regression specified as

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \dots + \beta_q x_q \qquad (1)$$

Where $X_1, X_2 \dots, X_q$ are the independent variable, Y_i is a dependent variable of the study, β_0 is the intercept while $\beta_{1,} \dots, \beta_q$ are the regression coefficients associated with each variable of interest.

$$P_i = \frac{1}{1+e^{-(\beta_0 + \beta_1 x_1 + \beta_1 x_2 + \dots + \beta_q x_q)}} \qquad (2)$$

Which is called a logistic response function for any value of $x_1, \dots, x_q$.The right-hand side of equation (3) will always lead to value in the interval (0, 1).

Sometimes it is preferable to utilize a different measure of belonging to a different class. The commonest is the measures know as odds. The odds belonging to class 1 is defined as the relation of the probability of belonging to class 1 to the probability of belonging to class 0. That is

$$\text{Odds} = \frac{p}{1-p} \qquad (3)$$

Again, for any given odds, the computation of the probability is defined as:

$$P = \frac{odds}{1+odds} \qquad (4)$$

So that the relationship between the odds and the probability is corrected by

$$odd = e^{\beta_0 + \beta_1 x_1 + \beta_1 x_2 + \dots + \beta_q x_q} \qquad (5)$$

Taking a log on both sides, we get the standard formulation of a logistic Model given as follows:

$$log(odds) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \ldots + \beta_q x_q \quad (6)$$

and let

$$Y_i = log\frac{p}{1-p} \quad (7)$$

Therefore, a logistic model with three categories has two logit functions.

(i) Logit function y=1 relative to logit function y=0

(ii) Logit function for y=2 relative to logit function y=0

Category y=0 is called a reference group

Equation (7) is now expressed as

$Log\big(g(1)\big) = \beta_i + \beta_{i1} X_i + \ldots + \beta_{1k} X_k$

$Log\big(g(2)\big) = \beta_2 + \beta_{i2} X_i + \ldots + \beta_{1k} X_k$

$Log\big(g(3)\big) = Log 1 = 0$

Therefore, in multinomial logistic regression, we have the following

$$f(1)\frac{g(1)}{g(1)+g(2)+1} = \frac{e^{\beta_i + \beta_{i1} X_i + \ldots + \beta_{1k} X_k}}{e^{\beta_i + \beta_{i1} X_i + \ldots + \beta_{1k} X_k} + e^{\beta_2 + \beta_{i2} X_i + \ldots + \beta_{1k} X_k} + 1} \quad (8)$$

$$f(2)\frac{g(2)}{g(1)+g(2)+1} = \frac{e^{\beta_2 + \beta_{2i} X_i + \ldots + \beta_{2k} X_k}}{e^{\beta_i + \beta_{i1} X_i + \ldots + \beta_{1k} X_k} + e^{\beta_2 + \beta_{2i} X_i + \ldots + \beta_{2k} X_k} + 1} \quad (9)$$

$$f(3) = \frac{1}{e^{\beta_i + \beta_{i1} X_i + \beta_{1k} X_k} + e^{\beta_2 + \beta_{i2} X_i + \beta_{2k} X_k} + 1} \quad (10)$$

Here we assume f (1) to the probability of the respondent experiencing domestic violence Always f (2) is the probability that the respondent has never experienced domestic violence and f (3) for the respondent experiencing domestic violence sometimes. All statistical analyses were conducted using SPSS version 25. Initial data exploration included descriptive statistics (frequencies, percentages, means, and standard deviations) to characterize the sample. The core inferential analysis involved Multinomial Logistic Regression (MLR) to assess the relative likelihood of experiencing each form of violence. The MLR model was specified with emotional violence as the reference category, allowing comparisons between emotional vs. physical and emotional vs. sexual violence.

All independent variables were entered simultaneously to account for confounding and interaction effects. Model fit was evaluated using:

- Likelihood Ratio Tests
- Pseudo R^2 measures (Nagelkerke and Cox & Snell)
- Classification accuracy
- The significance level was set at $p < 0.05$.

This analytic approach allowed for the simultaneous estimation of multiple logit functions and provided insight into the unique predictors of each violence type, in line with the modeling requirements of unordered categorical data. MLR was used to estimate the log odds of physical and sexual violence relative to emotional violence, following standard logistic regression frameworks (Hosmer et al., 2013; Long & Freese, 2006). Model fit was evaluated via Likelihood Ratio Tests, Nagelkerke R^2, and classification accuracy (Saha, 2011; McCullagh & Nelder, 1989).

The study obtained ethical clearance from the Plateau State Specialist Hospital Research Ethics Committee. All participants provided written informed consent after being briefed on the study objectives, procedures, and confidentiality safeguards. Given the sensitivity of the subject matter, respondents were assured of anonymity, and those in distress were referred to the hospital's psychosocial support unit for counseling and follow-up care. The study analyzed data from 499 pregnant women attending antenatal care at Plateau State Specialist Hospital, Jos. The distribution of domestic violence forms reported by respondents was as follows:

Emotional violence: 41.5%

Physical violence: 28.7%

Sexual violence: 18.2%

No violence: 11.6% (excluded from MLR analysis to focus on nominal outcomes)

The mean age of respondents was 28.4 years (SD ± 6.1). A majority (62.8%) had attained at least secondary education. Most were either self-employed or unemployed, and partner substance use was reported by 34.1% of the respondents. These background variables were included as covariates in the multinomial regression model to determine their association with the type of domestic violence experienced.

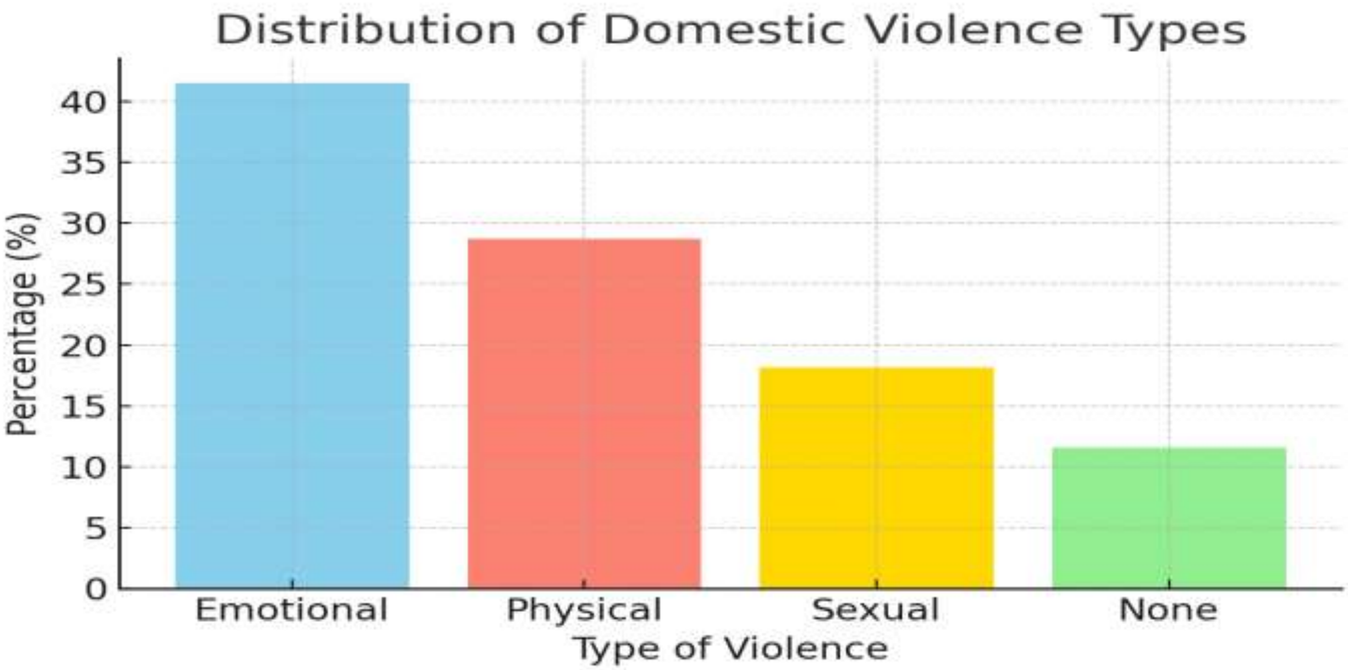

Figure 1: Distribution of reported domestic violence forms among respondents.

Results

To identify the determinants of the form of domestic violence experienced, a Multinomial Logistic Regression (MLR) model was estimated, using emotional violence as the reference category. This allowed the simultaneous comparison of risk factors associated with physical and sexual violence relative to emotional violence. Emotional violence was the most common (41.5%), followed by physical (28.7%) and sexual (18.2%) violence. Cases reporting no violence (11.6%) were excluded from the MLR analysis.

Predictors of Physical Violence (vs Emotional Violence)

- Low maternal education ($p = 0.004$)
- Partner substance use ($p < 0.001$)
- History of mental health issues ($p = 0.029$)

Predictors of Sexual Violence (vs Emotional Violence)

- Unemployment ($p = 0.017$)

- Younger maternal age ($p = 0.038$)
- Partner substance use ($p = 0.001$)

Model Fit and Classification Accuracy

- Nagelkerke $R^2 = 0.39$
- Likelihood Ratio $\chi^2 = 152.8$, df = 24, $p < 0.001$
- Model correctly classified 73.2% of cases

Summary of Key Findings

The MLR model revealed distinct predictor profiles for physical and sexual violence when compared to emotional violence:

- Partner's substance use was a strong, consistent risk factor across all forms of violence.
- Socioeconomic vulnerability—reflected in low education and unemployment—differentiated physical and sexual violence risks.
- Maternal mental health history and age further stratified exposure risks, underscoring the need for multidimensional screening in antenatal care settings.

Model Performance

- Nagelkerke $R^2 = .39$
- Likelihood Ratio = 152.8, df = 24, $p < .001$
- Classification accuracy: 73.2%

Discussion

MLR enabled the examination of distinct risk profiles for different forms of domestic violence among pregnant women. Emotional abuse emerged as the most prevalent, aligning with findings from similar Nigerian studies (Ashimi & Amole, 2015; Bashir & Bello, 2020). Partner substance use was a strong and consistent predictor of both physical and sexual violence, confirming its central role in driving violent behavior (Campbell, 2002; WHO, 2013). Socioeconomic factors such as low maternal education and unemployment increased the likelihood of abuse, in line with the ecological and social learning theories of domestic violence (Bandura, 1977; Bronfenbrenner, 1979). Younger age also predicted sexual violence, suggesting a compounded vulnerability among younger expectant mothers. Methodologically, MLR proved more appropriate than binary models, as it allowed for simultaneous estimation of multiple non-ordinal outcomes (Agresti, 2018; Long & Freese, 2006). Model fit metrics confirmed the robustness and predictive validity of the model, supporting its use in health outcomes research involving categorical data (Hosmer et al., 2013).

Conclusion

This study reaffirms the utility of MLR in modeling unordered categorical outcomes in public health (Agresti, 2018; Saha, 2011). It highlights the prevalence of domestic violence during pregnancy and the multifactorial nature of its determinants. Policy and clinical strategies should incorporate psychosocial screening, partner-focused interventions, and economic empowerment to address these risks effectively (Tjaden et al., 2000; Stark, 2016).

Recommendations

Policy: Integrate routine screening for emotional, physical, and sexual forms of domestic violence into national antenatal care protocols to enhance early detection and intervention (WHO, 2013).

Practice: Strengthen the capacity of healthcare providers through targeted training on the identification, documentation, and referral of domestic violence cases within maternal health services (Barnett et al., 2011; Bronfenbrenner, 1979)

Research: Future investigations should adopt longitudinal study designs to capture causal dynamics and examine mediating variables such as social support networks, mental health status, and coping strategies.

Limitations

- Cross-sectional design limits causal inference.
- Self-reported data may be subject to underreporting due to stigma.
- Results are facility-based and may not generalize to community settings.

References

Agresti, A. (2018). Statistical methods for the social sciences (5th ed.). Pearson.

Ashimi, A., & Amole, T. (2015). Prevalence and predictors for domestic violence among pregnant women in a rural community, Northwest Nigeria. Nigerian Medical Journal, 56(2), 118. https://doi.org/10.4103/0300-1652.150696

Bandura, A. (1977). Social learning theory. Prentice-Hall.

Barnett, O. W., Miller-Perrin, C. L., & Perrin, R. D. (2011). Family violence across the lifespan: An introduction (3rd ed.). Sage.

Bashir, A., & Bello, M. (2020). Prevalence and determinants of domestic violence among pregnant women in Northwestern Nigeria. Journal of Public Health in Africa, 11(1), Article 1092.

Bronfenbrenner, U. (1979). The ecology of human development. Harvard University Press.

Campbell, J. C. (2002). Health consequences of intimate partner violence. The Lancet, 359(9314), 1331–1336. https://doi.org/10.1016/S0140-6736 (02)08336-8

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (3rd ed.). Wiley.

Iliyasu, Z., Abubakar, I. S., Galadanci, H. S., & Aliyu, M. H. (2013). Intimate partner violence among pregnant women in Kano, Northern Nigeria. Journal of Interpersonal Violence, 28(4), 812–827. https://doi.org/10.1177/0886260512455861

Long, J. S., & Freese, J. (2006). Regression models for categorical dependent variables using Stata (2nd ed.). Stata Press.

McCullagh, P., & Nelder, J. A. (1989). Generalized linear models (2nd ed.). Chapman & Hall.

Rabin, R. F., Jennings, J. M., Campbell, J. C., & Bair-Merritt, M. H. (2009). Intimate partner violence screening tools: A systematic review. American Journal of Preventive Medicine, 36(5), 374–385. https://doi.org/10.1016/j.amepre.2009.01.024

Saha, S. (2011). Binary logistic regression model for analysing school examination results. International Journal of Statistics and Systems, 6(2), 147–156.

Seligman, M. E. P. (1975). Helplessness: On depression, development, and death. Freeman.

Stark, E. (2016). Coercive control. Oxford University Press. https://doi.org/10.1093/oso/9780197639986.001.0001

Tjaden, P., Thoennes, N., & Centers for Disease Control and Prevention. (2000). Extent, nature, and consequences of intimate partner violence. National Institute of Justice.

World Health Organization. (2013). Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. WHO Press.