ORIGINAL ARTICLE

Modelling the Number of Household Visit to Health Care Centres in Some Nigeria Communities Using Count Data Regression Models

Samuel Olorunfemi Adams¹,*, Davies Abiodun Obaromi², Rauf Ibrahim Rauf¹

¹Department of Statistics, University of Abuja, Abuja, Nigeria.
²Department of Statistics, Confluence University of Science and Technology, Lokoja, Kogi State, Nigeria.

ARTICLE INFO

ABSTRACT

Introduction: The need to model the impact of some demographic indicators on the frequency of household visits to healthcare centres in Nigeria's community is very important for preventing and spreading community diseases. This study aimed to investigate the effect of the patients' age, gender, marital status, type of illness and amount spent on the frequency of visits to community health care centres in Nigeria and to compared Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR) models to determine the preferred count regression model for the number of household visits to health centres in some communities in Nigeria.

Methods: Survey of 132640 households in some Nigeria communities obtained from the 2018/2019 Nigeria Living Standard Survey (NLSS) were extracted from the National Bureau of Statistics (NBS) in collaboration with the World Bank. The Negative Binomial and Generalised Poisson regression models were used to investigate the five demographic variables on the frequency of visit to the community health centres. The performance of the count regression model was assessed using the Chi-square -2log Likelihood Statistic (2logL), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) selection criteria.

Results: Findings from the study showed that the type of illness and amount spent has a significantly positive effect on the number of household members' visits to the community health care centres in Nigeria while age, gender, and marital status was discovered to have a negative effect on the number of household members' visits to the community health care centres in Nigeria.

Conclusion: The Nigeria Government, health centre management and community healthcare service providers' need to be aware that the amount spent and the nature of illness determines the level of health care services utilisation in the Nigeria community, hence the need for the drastic reduction in charges to encourage a household visit to the community health centres whenever the need arises.

Introduction

The visit of household members to community health centres is very important for access to health services through facilities located in their community to prevent the spread of community diseases (1). Community health care centre forms an integral part of the countries' health system.
It is the central function and the main focus of the community's overall social and economic development. It is also the first level of contact of individuals, families and communities with the national health system, bringing healthcare as close as possible to where people live and work and constitutes the first element of continuing health care process (2). The inception of community health care centre has facilitated some communities to have at least one health care facility sited as close as possible to where they live or work in all the districts or local government areas of many of the states in Nigeria. However, citing health care facilities does not necessarily translate to its utilisation; more so, one of the major factors maintaining high mortality rate in Nigeria is poor access to and utilising health services (3). The health centres in Nigeria communities face numerous problems like increased infectious and viral diseases like malaria, HIV/AIDS, and Corona Virus. Other challenges include; poor hygiene, corruption, malnutrition, lack of access to safe drinking water, poor health infrastructure, fake drugs, insufficient financial investment, lack of sufficient health personnel and an ongoing strike by health workers. The government's performance in the health sector has been abysmal (4). Investment in infrastructure has been poor, and meagre remuneration for health workers has created a massive brain drain to the United States of America and Europe. The government’s annual budget for the health sector is 4.17% of the total national budget, equivalent to only $5 per person per year (4). The essential medical care on which the Nigerian medical services framework effectively solves the numerous health problems in Nigeria (5, 6) additionally, it indicated that the quantity of revealed jungle fever cases expanded from 2,834,174 out of 2008 to 4,295,689 in 2009.4 National wellbeing frameworks stay frail while its administration is incapable and wasteful. There is no fair dispersion of human resources among the metropolitan and rustic region. More than 70 per cent of doctors is in urban areas where only 48 per cent of the population live, leaving 52 per cent of the people who live in the rural area at the mercy of inadequate health personnel. Research on the frequency of household visit to community health care centres shows that; Household power dynamics and economic considerations (poverty-related factors) were identified as the prevalence and characteristics associated with a visit to health care services in a rural area of Kenya (7). Findings from the study by (8) indicated that education, household wealth, and urban-rural residence are the most significant and consistence effects of women from Kenya, Tanzania, Nigeria, Pakistan, Bangladesh and India visit to health services for delivery. (9) studied the frequency of household visit to the Health Extension Programme (HEP) in Ethiopia. Their finding showed that HEP had a significant positive association with HEP implementation-related factors, such as the number of years after graduation, frequency of household visits by Health Extension Workers, and understanding the HEP. (10) suggested that the frequency of women's visit to maternal health centre can be improved through interventions targeted at individuals, household and community. (11) discovered that; income source, transportation, and health literacy were the community perception of the factors influencing healthcare visit in Ugandan. (12) showed that the most important factors influencing frequent pregnant women's visit to maternal health services (Antenatal and delivery services) in Ethiopia are; demographic,
socioeconomic and health-related factors. The pattern of the visit to child healthcare services provided by the Thana Health Complex (THC) of Keraniganj suggested that socioeconomic and poverty status exerts a dominating role in their children visiting child healthcare services (13, 14) discovered that characteristics of the health delivery system, social structure, and features of the individual affect the frequency of visit to maternity health centres in Nepal. (15) examined the frequency of visits to community clinics for diabetic care services. It was discovered that the determinants of frequent community visits in diabetes patients include urban residence, lower household income, lack of health insurance, lack of telephone follow-up and lack of household visit services. Socio-economic status and employment status were found to be strongly related to access and utilization of Zimbabwe’s health facilities (16).

A summary of the studies on community household visit to health care centres' by scholar shows that application of regression count data models to the frequency of visits to community health care centres in Nigeria have not been investigated so far. The gap in existing literature shall be filled in this study. This study aims to examine the effect of the patents' age, gender, marital status, type of illness and amount spent on the frequency of visits to community health care centres in Nigeria and to compared NBR and GPR models to determine the preferred count regression model for the number of household visits to health centres in some communities in Nigeria.

In the first section, the background information on health care was discussed. Second, the method and source of data collection were examined, the three count data regression methods were presented, and the goodness-of-fit criteria were evaluated. Third, the results computed were presented and discussed. The study was concluded with a discussion of the finding, implications, and suggestion of future household visits to Nigeria's community health care centres.

**Materials and methods**

**2.1 Source of data and method of data analysis**

The data utilised in the study is a secondary data obtained from Living Standard Survey (NLSS) data collected by the Nigeria Bureau of Statistics (NBS) in collaboration with the World Bank and Department for International Development (DFID) for the year 2018/2019 (17). Data on the number of visits to the health care centres of some selected communities in Nigeria is the dependent variable. The demographic information data that include gender, age, marital status, type of illness and amount spent in the health care centres are independent variables. The data is processed into a health expenditure table, consisting of expenditure and the demographic data variables. The information was tested for Poisson distribution using the one-sample Kolmogorov-Smirnov test, followed by the summary statistics for the discrete and categorical variables for all independent variables. The Omnibus test for the overall significance of the independent variables was performed for the number of visits to the community health centres; this was followed by the over-dispersion test using the Pearson chi-square and Deviance. These determined whether the data adhered to a poisson regression, negative binomial or the generalised poisson regression models. The Poisson regression was dropped in favour of the negative binomial and generalized poisson regression because the data
possessed over-dispersion. All computation was carried out using STATA version 16.

2.2 Poisson Regression (PR)

Poisson regression contains overdispersion (variance is greater than the mean value's value) (18). Let \( Y \) be a random variable (the number of visits to community health care centres) and let be the outcomes of the cases be an event. The variable \( Y \) is said to follow a Poisson distribution with parameter \( \lambda > 0 \) if the probability function is given by:

\[
p(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \tag{1}\]

Where \( n = 1, 2, 3, 4 \) and 5, is the number of occurrences of an event and \( \lambda \) is defined

- \( X_1 = \text{age} \)
- \( X_2 = \text{gender} \)
- \( X_3 = \text{Marital status} \)
- \( X_4 = \text{type of illness} \)
- \( X_5 = \text{amount spent in the health care centre} \)

\[
\log(\theta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \tag{3}\]

The parameter \( \beta \) represents the expected change in the logarithm of the mean per unit change in the predictor \( X_i \) which can be estimated by the Maximum Likelihood Estimator (MLE) method.

2.3 Generalised Poisson Regression (GPR)

Generalised Poisson Regression (GPR) was developed to handle the Poisson regression model's equi-dispersion violations. It is mainly used to fit the overdispersed model, for situations where \( \text{Var}(y_i) > E(y_i) \) as well as under dispersion, \( \text{Var}(y_i) < E(y_i) \) (19). GPR model with parameter \((\mu, \theta)\) However, the Poisson regression models expected that the segments are haphazardly randomly distributed to general Poisson. In the investigation of GPR, if \( \theta \) is equivalent to 0, the model will be the model Poisson. If \( \theta \) is more than 0, GPR models address information containing tally overdispersion case and if \( \theta \) is under 0, address information containing under scattering tally. It was suggested that when \( y_i \) are a count response variable and it follows a Generalised Poisson
distribution, the probability density function given that \( i = 1, 2, \ldots, n \), then;

\[
f(y_i, \mu_i) = \left[ \frac{\mu_i}{1 + a \mu_i} \right]^{y_i(1+ay_i)\gamma_i^{-1}} \exp \left[ \frac{V_i(1 + ay_i)}{1 + a \mu_i} \right], y_i = 0, 1, 2, \ldots
\]

Where mean is given as, \( \mu_i = E(y_i) \) and variances \( \text{var}(y_i) = (\mu + a \mu_i)^2 \) and \( \mu_i \) is referred to as the dispersion parameter (20) and (21). Generalised Poisson distribution is a common increase in Poisson distribution (22). When \( \alpha = 0 \), the model in equation (1) reduces to the Poisson (2), where; \( \text{Var}(y_i) > E(y_i) \). When \( \alpha > 0 \), it means the variance \( \text{Var}(y_i) \) of the distribution represents count data with over-dispersion if \( \alpha > 0 \), it means the variance is less than the expectation, \( \text{Var}(y_i) < E(y_i) \), which simply means that the distribution represents count data with under-dispersion.

\[2.4 \text{ Negative Binomial Regression (NBR)}\]

Negative Binomial distribution is used to dispense the problem of over-dispersion in

\[
p_r[y_i = y] = \sum p_r(y_i = y/\theta_i) f(\theta_i) d\theta_i = \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{\Gamma\left(y_i + 1\right)\Gamma\left(\frac{1}{\alpha}\right)} \left(\frac{1}{1 + a \mu_i}\right)^{\frac{1}{\alpha}} \left[\frac{a \mu_i}{1 + a \mu_i}\right]^{y_i}, \text{for } i = 0, 1, 2, \ldots
\]

Where mean is given as,

\[
\mu_i = E(y_i) = V_i\left[e^{\Sigma_{\alpha \beta} \alpha_{ij} \beta_i} \right], \text{for } i = 1, 2, 3, \ldots, n \text{ while the variance of } y_i \text{ is given as; } \text{var}(y_i) = (\mu + a \mu_i)^2
\]

Where \( \alpha > 0 \), the model would be referred to as a dispersion parameter, the Poisson regression model can be regarded as a limiting model of the Negative Binomial Regression model as \( \alpha \) approaches 0.

\[2.5 \text{ Goodness-of-fit and model Selection Criteria}\]

When many regression models are available for a given data set, one can compare alternative models based on some goodness-of-fit measures. Several goodness-of-fit measures have been proposed in the frequency of household visits to health cares in the communities. The Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Chi-Square -2log likelihood model selection criteria are commonly used measures (24) and (25).

**Results**

As observed from this result, the variable exhibits over-dispersion, a situation where the standard deviation is higher than the mean. Over-dispersion in the Poisson
regression model may underestimate the standard error and significance of regression parameters. This condition leads to misleading inference about the regression parameters.

3.1 Test for overall significance of the explanatory variables

The Omnibus test result is presented in table 1. A likelihood ratio test of whether the independent variables, age group, gender, marital status, type of illness and amount spent in the health care centres collectively improve the model over the intercept only model. The result from the table indicates that all the independent variables have a p-value of .000 (i.e., p < .05), indicating a statistically significant overall model.

3.2 Over-dispersion test

In the Poisson regression modelling, there is an assumption about equidispersion that the mean value and the variance must be equal. However, this assumption is rarely met because overdispersion is present in most cases. To detect overdispersion, the value of Deviance when divided by the degree of freedom must be greater than one but when the division of deviance value by degree of freedom (d.f.) is less than one, then there exist under dispersion. Table 2 shows that the value of deviance/df and Pearson chi-square/df is greater than one. It can be deduced on the Poisson Regression models that the number of visits to community health care centre in Nigeria occurred in overdispersion. The over-dispersion problem is overcome by modelling with the Negative Binomial Regression (NBR) and Generalised Binomial Regression (GBR) because both methods can accommodate the dispersion parameter.

| Table 1: Omnibus test and goodness-of-fit test of the number of visits to health care centres |
|-----------------------------------------------|--------|------|---------|--------|-----------|
| Count Regression models                        | LR Chi-square | d.f. | p-value | 2LogL  | AIC       | BIC       |
| Poisson regression                             | 325.6       | 15   | .000    |        |           |           |
| Negative Binomial regression                   | 108.6       | 15   | .000    | -1472.312* | 2974.623* | 3038.913* |
| Generalized Poisson regression                 | 130.3       | 15   | .000    | -2241.598 | 4513.197  | 4577.487  |

* Refers to the best-fitted model, i.e. Negative Binomial Regression

| Table 2: Deviance and Pearson Chi-Square test of Poisson Regression |
|---------------------------------------------------------------|--------|------|---------|
| Criteria                        | Value     | DF.  | Value/D.F |
| Deviance                        | 2323.328  | 521  | 4.459    |
| Pearson Chi-Square              | 3131.666  | 521  | 6.011    |

The parameter estimation of the Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR) model presented in Table 3 shows that all variables under the type of illness and amount spent in the health care centres by the household members are significant at 5% level. It can be seen that the p-values of all parameters were less than...
0.05. The result also revealed that age group, gender and marital status were not significant at the 0.05 level of significance because their p > 0.05. The type of illness and amount spent by household members at the health care centre positively affected the household visit to health care centres. In contrast, age group, gender and marital status negatively affected the number of visits. The result is confirmed in estimating the predictors using the Negative Binomial Regression (NBR) model, a 1% increase in illness, injury, general checkup, pre and post-natal, amount spent in the health care centre, the number of males, married (monogamy), divorced and separated a will lead to 3.726, 3.501, 3.903, .036, .135, .584 and .361 increase in the number of days the household members spent in the healthcare centres in Nigeria. Also, the z-statistics values of 5.485, 5.369, 5.580 and 7.175 and 3.69 for illness, injury, general checkup, pre and post-natal and amount spent in the health care centres respectively are greater than two and by the rule of the thumb, thus confirming that they have a significant positive effect on the number of visits to health care centres in Nigeria.

However, a unit increase in age groups 0-39 years, 40-59 years, 60-79 years, informal and widowed will lead to -.223, -.134, -.207, -.211 and -1.109 decrease in the number of days the household members visit the healthcare centres in Nigeria. The z-statistic coefficient of .715, .279, .605, 1.191, 1.126, 1.414, .222, .903 for 0-39 years, 40-59 years, 60-79 years, male, married (monogamy), married (polygamy), informal and widowed shows that the number of visit to health care centres reduces because the values are less than two and by the rule of the thumb, it implies that age groups 0-39 years, 40-59 years, 60-79 years, informal and widowed have a negative effect of the number of visits to the health care centres.

The parameter estimation of the Generalised Poisson Regression (GPR) model presented in Table 3 shows that all variables under the type of illness, amount spent, and age group of the household members visit the hospital are significant at a 5% level. It was discovered that the p-values of all parameters were less than 0.05. The result also revealed that gender and marital status were not significant at the 0.05 level because p > 0.05. The male gender, married (monogamy), divorced, separated, illness, injury, general checkup, pre and post-natal and amount spent had a positive effect on the household visit to health care centres while married (polygamy), informal marriage and widowed had a negative impact on the number of visits.

Estimating predictors using the Generalised Poisson Regression (GPR) model shows that, if the number of male household members, married (monogamy), divorced, separated, illness, injury, general checkup, pre and post-natal and amount spent are increased by a unit, the frequency of household visit to the community health care centre will also increase by .115, .177, .640, .435, 1.677, 3.867, 1.923, 1.270 and .048 respectively. However, a unit increase in age groups 0-39 years, 40-59 years, 60-79 years, married (polygamy), informal and widowed will lead to -.392, -.371, -.455, -.270, -.1.109 and -.221 decrease in the number of household members' visit to the healthcare centres in Nigeria.

Table 3 also shows the goodness of fit criteria results for the two count data regression model, namely, Negative Binomial Regression (NBR) and Generalized Poisson Regression. The result indicated that the Negative Binomial Regression (NBR) has the least AIC (2974.623), BIC (3038.913), and -
2log likelihood (-2241.598). This result indicates that NBR is the preferred count data regression model for modelling age group, gender, marital status, type of illness and amount spent at the health centres on the number of household visits to health care centres in some Nigeria communities.

Table 3: Parameter Estimates of Negative Binomial and Generalised Poisson Regression

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Negative Binomial</th>
<th>Generalised Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coefficient</td>
<td>P&gt;</td>
</tr>
<tr>
<td>Intercept</td>
<td>-.386</td>
<td>.790</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-39 years</td>
<td>-.223</td>
<td>.398</td>
</tr>
<tr>
<td>40-59 years</td>
<td>-.134</td>
<td>.597</td>
</tr>
<tr>
<td>60-79 years</td>
<td>-.207</td>
<td>.437</td>
</tr>
<tr>
<td>80 years and above</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.135</td>
<td>.275</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Marital Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married (monogamy)</td>
<td>.136</td>
<td>.275</td>
</tr>
<tr>
<td>Married (polygamy)</td>
<td>-.211</td>
<td>.289</td>
</tr>
<tr>
<td>Divorced</td>
<td>.584</td>
<td>.235</td>
</tr>
<tr>
<td>Separated</td>
<td>.361</td>
<td>.638</td>
</tr>
<tr>
<td>Widowed</td>
<td>-.167</td>
<td>.342</td>
</tr>
<tr>
<td>Single</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Type of illness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Illness</td>
<td>3.726</td>
<td>.002</td>
</tr>
<tr>
<td>Injury</td>
<td>3.501</td>
<td>.021</td>
</tr>
<tr>
<td>General checkup</td>
<td>3.903</td>
<td>.020</td>
</tr>
<tr>
<td>Pre &amp; post-natal</td>
<td>3.262</td>
<td>.019</td>
</tr>
<tr>
<td>Childbirth</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Amount spent</td>
<td>.036</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Discussion

The novelty of this study is that; the type of illness and amount spent in the health care centre has a significant positive effect on the number of household members' visits to the community health care centres in Nigeria. In Bakeera et al. (2009) study, income source, transportation, and health literacy significantly affected the community perception of the factors influencing healthcare visit in Ugandan. Kevany et al. (2012), in a study on the socioeconomic status and health care utilisation in rural Zimbabwe, discovered that socioeconomic status and employment status had a strong association with health care visit in rural Zimbabwe.

Our result also indicated that a 1% increase in illness, injury, general checkup, pre and post-natal, amount spent in the health care centre, the number of males, married (monogamy), divorced and separated will lead to 3.726, 3.501, 3.903, .036, and .135, .584 respectively increase in the number of days the household members spent in the healthcare centres in Nigeria. Jimma et al. (2011), in a study on the utilisation of health services in Ethiopia, discovered that gender (0.23), marital status (8.1), household income (0.7), socioeconomic status (3.5), presence of an illness (28.3), perceived transport cost (0.15) and distance to the nearest health centre or hospital (2.9) were found to be predictors of health care visit. These results were found to be slightly higher than our study results.

The other demographic information of the household members like age, gender, and marital status was discovered to negatively impact the number of household members' visits to the community health care centres in Nigeria. A unit increase in age groups; 0-39 years, 40-59 years, 60-79 years, informal and widowed resulted in -.223, -.134, -.207, -.211 -1.109 decrease in the number of days the household members visit the healthcare centres in Nigeria. This result was somehow similarly to the result from Develay et al. (1996), where; age, socioeconomic level, illness characteristics and cost of care and transportation were the found to be the determinants of frequency of visits to modern health care services in Ouagadougou, Burkina-Faso.

Conclusion

In this study, two count data regression technique were employed to model the effects of age, gender, marital status, illness type and amount spent on treating illness on the number of visits to health care centres in some selected communities in Nigeria. The data used in the study is secondary data obtained from Living Standard Survey (NLSS) data collected by the Nigeria Bureau of Statistics (NBS) in collaboration with the World Bank and Department for International Development (DFID) for the year 2018/2019. Data on the number of visits to the health care centres of some selected communities in Nigeria is the dependent variable. The demographic information data that include gender, age, marital status, type of illness and amount spent in the health care centres are independent variables. The results obtained from the analysis revealed; the type of illness and amount spent has a significantly positive effect on the number of household members' visits to the community health care centres in Nigeria while demographic indicators like; age, gender, and marital status was discovered to have a negative effect on the number of household members' visits to the community health care centres in Nigeria. Findings from the study
also indicated that; Poisson Regression was not suitable for modelling count data with over-dispersion, so other count data regression techniques like; Negative Binomial Regression (NBR) and Generalized Poisson Regression (GPR) were applied to the count data. The result shows that the best model for the number of visits to the community's health care centre is obtained from the negative binomial regression model; this implies that the negative binomial regression model is more appropriate for modelling count data over-dispersion. It was inferred from the results generated from the model selection criteria applied, which include; -2logL = -1472.312, AIC=2974.623, and BIC=3038.913. The three measures established that Negative Binomial Regression as the best model because it had the smallest value of all three selection criteria.

Conflict of Interests
We have no conflicts of interest to disclose.

Acknowledgements
Authors are thankful to the Department of Statistics, Faculty of Science, University of Abuja, Nigeria, for their support and encouragement throughout the period of writing this work.

References
