APPLICATION OF ZERO-INFLATED POISSON MODEL TO DISABILITY IN TYPE 2 DIABETIC DATA.

[1]Oritogun, Kolawole S., [2]Oyewole, Olufemi O.
[1]Department of Community Medicine and Primary Care, Faculty of Clinical Sciences, Obafemi Awolowo College of Health Sciences, Olabisi Onabanjo University, Sagamu, Ogun State.
[2]Department of Physiotherapy, Olabisi Onabanjo University Teaching Hospital, Sagamu, Ogun State.

Corresponding Author: Oritogun Kolawole S. E-mail:oritogunks@yahoo.co.uk, ksoritogun@gmail.com

ABSTRACT

Disability in type – 2 diabetic data is a discrete quantitative data which is mostly modeled by the Poisson regression. This type of data sometimes has the variance greater than the mean which is called over-dispersion; and Poisson regression may not be able to give significant estimation to the data. Therefore, this study aimed at identifying appropriate model to analyze type 2 diabetic patients' disability. Methods: This was cross-sectional data of patients with type 2 diabetic attending a University Teaching Hospital in South-West Nigeria. Disability was the outcome variable while explanatory variables were: age, sex, marital status, disability duration, education, occupation, diastolic blood pressure, fasting blood glucose, body mass index, and glycosylated haemoglobin ($HBA1_C$). Poisson and Zero-Inflated Poisson (ZIP) models were applied to the data and compared. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to select a better model. P-value of 5% was considered to be significant. Results: ZIP model had the smaller AIC and BIC of 2349.27 and 2407.82 respectively. Therefore, ZIP model seems to be a better model. Age (IRR = 0.997, $p < 0.05$), marital status (single, IRR= 1.31, $p<0.01$), BMI (IRR= 0.991, $p<0.01$), Teaching occupation (IRR= 1.705, $p<0.01$) were some of the associated factors of disability. Conclusion: ZIP model was the better model in the presence of excess zeros in identifying factors associated with disability in type 2 diabetic patients.

Key words: Type 2 diabetic, Disability, Zero-Inflated Poisson, AIC, BIC,

Accepted Date:

Introduction

One of the oldest diseases affecting humans is diabetes mellitus (DM) which is a public health challenge among others. Public Health problems which result in chronic complications such as kidney damage, blindness affecting individuals if proper management is lacking . Serious attention is needed since the prevalence of disability continues to increase worldwide.

Generally, prevalence of type 2 diabetes mellitus (T2DM) is more in the older adults and adult males. The prediction was that the prevalence of T2DM will increase especially in developing countries like Nigeria in the next one decade . There were variations, in the geographical areas, education and income, and being rural or urban especially in Africa . In Australia, it was more among male (53%) than female (47%), diabetes is common among older people than the younger people though this report is conservation because some people cannot say whether they have it or not and even if they have, the time of commencement cannot be ascertained. Diabetes was more pronounced among people with lower education than those who had university degree which was attributed to the fact that those with lower education expose themselves to risk factors such as smoking and alcohol consumption. It was reported that 12.9% of patients with diabetes engaged in active smoking with higher proportion among males, more in the Northern Africa than sub-Saharan. In Africa about 1% of the rural dwellers had diabetes compared with 5% to 7% in urban sub-Saharan.

In Nigeria it was estimated that about 10% of the population had diabetes (DM). The future of Nigeria as a nation regarding this diabetes depends on the increasing rate in population of Nigerians, increasing life expectancy, low per capital income of most Nigerians and many others .

From the above it is important to identify some likely factors that are associated with disabilities of diabetes patients using appropriate regression model. Many studies have been done which made use of one statistical method or the other to find out determinants of disabilities among diabetes patients. Oyewole *et al* made use of Poisson regression to identify predictors of Global disability burden (GDB) in Nigeria having fasting blood glucose, glycosylated haemaglobin, systolic blood pressure, DM duration, marital status and age to be significantly associated with GDB.

Some studies made use of descriptive statistics, chi-square tests of means to identify factors associated with disabilities in diabetes patients. Logistic regression model was used to assess factors associated with the quality of life among type 2 Diabetic patients in Korean .They found that age, Insulin use and family history were significantly associated with diabetic patients' quality of life. Differentials in gender and T2DM study made use of correlation for analysis of data. Obesity, physical activity, poor nutrition, genetic predisposition were some associated factors of T2DMas stated in a study from Netherlands. In addition, gender differentials and factors correlated with quality of life among people with T2DM were studied using Poisson model with different link functions by Oritogun and others. Moreover ,in South-Western Nigeria, health behavior and quality of life of T2DM were analyzed using descriptive statistics .Some outcome variables in T2DMare count data, for example, a disability in T2DM . Other studies from the literature whose outcome variables were count data include: Dental Carries indices with many zero counts, Cavies experience using Hurdle model and .

The common regression model for count data is Poisson regression model. This is characterized by equal means and variance but in a real life situation, this property may not hold. Then, it is necessary to apply a model that will take care of this deficiency. When the mean and variance are not equal, that means probably the variance is higher than the mean. Whenever the situation occurs, Negative Binomial (NBI) model, Zero-inflated models will be appropriate. Zero-inflated models arise when there are many zeros in the data. Therefore, the objective of this study was to identify an appropriate model for identifying risk factors of disability in type 2 diabetic patients. Proper identification of this model will give correct estimates and associated factors with global disability of T2DM.

METHODOLOGY

Settings & Subjects

The data used in this study was secondary data of patients attending outpatient clinic of a University Hospital in the South west zone of Nigeria. The sample size was 162 adults of age > 21 years recruited between March 2016 and May 2017. The assessment of disability of the patient was carried out using World Health Organization Disability Assessment Schedule (WHODAS 2.0).Details of the data were found in Oyewole and others (Oyewole, Odusan, and Ale 2019).

Ethical Consideration

This study employed secondary data and therefore, ethical approval was not required.

Outcome variables and Covariates

The outcome variable was global disability of the patients while the independent variables were age (X1) of the patient, sex (X2), marital status (X3), disability duration (X4), education (X5), occupation (X6), Diastolic blood pressure (DBP:X7), Fasting Blood Glucose (FBG:X8), Body Mass Index (BMI:X9) and glycosylated haemoglobin (HBA1C:X10).

Statistical Analysis

Statistical package for Social Sciences (SPSS) version 23 and R programming package were used for the data analysis. The frequency of the response variable (disability), mean and variance were determined. Poisson (PO) and Zero-inflated Poisson (ZIP) models were applied on the data; Incidence rate ratios (IRR) were given. Criteria for model selection were Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).P-value < 0.05 was considered to be significant.

Models (Javali & Pandit, 2010; Oritogun & Bamgboye, 2018; Pittman *et al.,* 2018)

1. The Poisson Model

Poisson model is useful for count data i.e. when the dependent variable is a count variable as can be seen in our study (disability outcome). The unique assumption of Poisson model is that the mean is equal to the variance.

Let a random variable Y be a count data for the i[th] subject; and X_i: be a vector of covariates. The Poisson distribution may be represented as:

$$P(x,\theta) = \frac{e^{-\theta}\theta^{x}}{x!}, x = 0,1,2,...,\theta > 0 \quad .... (2.1)$$

The mean and variance are equal showing E(X) = Var(X) = θ

Expressing parameter θ as a log linear model:

$$\log\theta_i = X^1\beta_i \quad ....(2.2)$$

Where β_i is the parameter.

$$\log\theta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_{10} X_{10} \quad .... (2.3)$$

2. Zero-inflated Poisson (ZIP) model

ZIP model is as follows

$$\Pr(Y_i / X_i) = \begin{cases} \rho_i + (1-\rho_i)\exp(-\mu_i) & for\ Y_i = 0 \quad .... (2.4) \\ (1-\rho_i)\exp(-\mu_i)\mu_i^{Y_i}/Y_i! & for\ Y_i > 0 \quad .... (2.5) \end{cases}$$

ρ_i represents the probability of the presence of extra zero.

Mean, E (Y) = $(1-\rho)\mu = \tau$

Variance, Var(Y)= $\tau + \left(\frac{\rho}{1-\rho}\right)\tau^2$

Model Selection (Yusuf, Afolabi, and Agbaje 2018)

The appropriate model for the set of data in this study was selected using two criteria namely: Akaike Information criterion (AIC) and Bayesian Information Criterion (BIC).

The AIC is given by the relationship

AIC = -2 log L (θ) + 2K

L (θ) = maximized likelihood function and

K= number of parameters (including the intercept). The model that has the minimum AIC value is the better or best model.

The BIC is given by the relationship

BIC = -2log L + P log (n)

P is the number of parameters in the model and n is the number of observations. The model with the minimum BIC is considered as the best model.

RESULTS

Poisson Regression Results

The mean, median and modal disability were 22.07, 19.00 and 0.0 respectively, while the variance of the response was 354.16. The percentage of zeros in the data was 14.2%.

Table 1 shows the estimates of Poisson regression model with incidence rate ratios (IRR). A unit increase in age of the respondents reduce the disability by 0.6% (IRR = 0.994, $p<0.01$). Respondents who were single had 56.6% higher disability than the married ones (IRR = 1.566. $p< 0.01$). Disability score was higher (10.4%) among respondents with Secondary education compared to respondents with Higher education (IRR = 1.104, $p < 0.05$), whereas respondents with no education experienced less disability (IRR = 0.693, $p <0.05$) than those with higher education. Disability among respondents in the teaching profession was 72% more than respondents who were Artisan. One unit increase in diastolic blood pressure increases disability by 0.4% ($p<0.05$). Also, one unit increase in Fasting blood glucose and Body Mass Index decrease disability by approximately 0.3% and 1.1% respectively ($p<0.001$). One unit increase in HBA1c increase disability of respondents by 5.4%.

Zero-Inflated Poisson Regression Results

The estimates and IRR of Zero Inflated Poisson Regression model are shown in Table 2. Each additional year of age is associated with an estimated decrease in disability (IRR= 0.997, $p<0.05$). Being single is 31% higher in disability that the married (IRR = 1.310, $p < 0.01$). Respondents with primary education had 19.1% decrease in disability compared to those with higher education (IRR = 0.809, $p < 0.01$). Respondents whose occupation was professional, teaching, trading, and other type of occupation were 32.8, 70.5, 32 and 73.3% respectively increased in disability compared to respondents who were artisan ($p < 0.01$).One unit increase in diastolic blood pressure and $HBA1_C$ of the respondents is associated with 0.6% and 3.9% increase respectively in disability ($p<0.01$). In addition, a unit increase in fasting blood glucose and Body Mass Index (BMI) decrease the disability of the respondents by 0.2% and 0.9% respectively ($p<0.01$).

Table 1: Estimates of Poisson Regression Model

Variable	Estimate	Std.Error	Z	IRR	p-value
Intercept	2.9241	0.2177	13.43	18.618	0.000
Age	-0.0055	0.0017	-3.24	0.994	0.001
Sex:					
Female+					
Male	-0.0518	0.0403	-1.29	0.950	0.198
Maritalstatus:					
Married+					
Separated/divorce	0.3867	0.1991	1.94	1.472	0.052
Single	0.4488	0.0796	5.64	1.566	0.000
Others	-0.0163	0.0520	-0.31	0.984	0.754
Disability duration	0.0005	0.0004	1.32	1.001	0.186
Education:					
Higher+					
Secondary	0.0990	0.0439	2.25	1.104	0.024
Primary	-0.0927	0.0512	-1.81	0.912	0.070
No eduaction	-0.3669	0.1573	-2.33	0.693	0.020
Occupation:					
Artisan+					
Professional	0.1140	0.0922	1.24	1.121	0.216
Teaching	0.5427	0.0912	5.95	1.721	0.000
Trading	0.1665	0.0854	1.95	1.181	0.051
Others	0.4596	0.0845	5.44	1.583	0.000
Diastolic Blood Pressure	0.0041	0.0016	2.51	1.004	0.012
Fasting Blood Glucose	-0.0026	0.0004	-6.05	0.997	0.000
Body Mass Index	-0.0109	0.0031	-3.52	0.989	0.000
HBA1C	0.0528	0.0054	9.80	1.054	0.000
AIC =3235.95 ; **BIC** = 3291.42					

Table 2: Estimates of Zero-inflated Poisson Regression Model

Variable	Estimate	Std.Error	Z	IRR	p-value
Intercept	2.7192	0.2225	12.22	15.168	0.000
Age	-0.0034	0.0017	-2.05	0.997	0.040
Sex:					
Female+					
Male	-0.0582	0.0397	-1.47	0.943	0.143
Marital status:					
Married+					
Separated/divorce	0.1379	0.1988	0.69	1.148	0.488
Single	0.2699	0.0786	3.44	1.310	0.001
Others	-0.0005	0.0519	-0.01	1.000	0.993
Disability duration	0.0007	0.0004	1.81	1.001	0.070
Education:					
Higher+					
Secondary	0.0305	0.0443	0.69	1.031	0.492
Primary	-0.2122	0.0508	-4.18	0.809	0.000
No eduaction	-0.1708	0.1582	-1.08	0.843	0.280
Occupation:					
Artisan+					
Professional	0.2838	0.0916	3.10	1.328	0.002
Teaching	0.5335	0.0905	5.90	1.705	0.000
Trading	0.2770	0.0846	3.28	1.319	0.001
Others	0.5500	0.0838	6.56	1.733	0.000
Diastolic Blood Pressure	0.0060	0.0016	3.73	1.006	0.000
Fasting Blood Glucose	-0.0018	0.0004	-3.97	0.998	0.000
Body Mass Index	-0.0088	0.0030	-2.89	0.991	0.004
HBA1C	0.0381	0.0056	6.85	1.039	0.000
AIC = 2349.27; **BIC**= 2407.82					

Model Comparison and Selection

Poisson regression model had AIC and BIC of 3235.95 and 3291.42 respectively while zero-inflated Poisson (ZIP) regression model had AIC and BIC of 2349.27 and 2407.82 respectively. ZIP model had smaller AIC and BIC. Therefore, ZIP was the better model to identify risk factors of disability in type 2 diabetic patients.

DISCUSSION

The estimates from the two models were not the same. Some of the estimates from Poisson model were smaller than zero-inflated Poisson, and some were bigger. Poisson model identified respondents with secondary education and no education being significantly associated with disability whereas, ZIP model did not. Professional and trading occupation were significantly associated with disability in ZIP which were not significant in PO model. The mean($\bar{x}$) median and mode were 22.07, 19.00 and 0.00 respectively. The variance, var (x) was 354.16 which shows over dispersion ($\text{var}(x) > \bar{x}$) (Famoye, Wulu, and Singh 2004; Pittman et al. 2018; Yusuf, Afolabi, and Agbaje 2018).

The model selection criteria have shown AIC (2349.27) and BIC (2407.82) for zero-inflated Poisson to be smaller than that of Poisson model. This shows that ZIP is considered to be a better model than PO. The frequencies and percentages of 0, 6, and 8 in disability were 23 (14.2), 16 (9.9) and 15 (9.3) respectively. These values indicated more or excess zeros in the values of disability. In this situation, ZIP model will be appropriate (Famoye, Wulu, and Singh 2004; Javali and Pandit 2010).

CONCLUSION

This study has shown that ZIP model was the better model to identify predicting factors of disability among type 2 diabetic patients where Poisson model assumptions of equal mean and variance could not hold. Also, ZIP is appropriate for data with excess zeros.

REFERENCES

Amballi, A A et al. 2019. "Cortisol and Electrolyte Levels in Type 2 Diabetic Patients Attending a State University Teaching Hospital, Ogun State, Nigeria." *International Journal of Biomedical and Health Sciences* 14(2): 51-58.

Amballi, Adebayo Adetola et al. 2018. "Prevalence of Microalbuminuria among Adults with Type 2 Diabetes Mellitus at OOUTH, Sagamu." *Annals of Health Research* 4(1): 15-21.

Australian Institute of Health and Welfare. 2013. *Diabetes and Disability*□

Chung, Jin Ook, Dong Hyeok Cho, Dong Jin Chung, and Min Young Chung. 2013. "Assessment of Factors Associated with the Quality of Life in Korean Type 2 Diabetic Patients." Internal Medicine 52(2): 179-85.

Colberg, Sheri R. *et al.* 2010. "Exercise and Type 2 Diabetes: The American College of Sports Medicine and the American Diabetes Association: *Joint Position Statement." Diabetes Care* 33(12).

Famoye, Felix, John T Wulu, and Karan P Singh. 2004. "On the Generalized Poisson Regression Model with An." 2: 287-95.

Hofstetter, Hedwig, Elise Dusseldorp, Achim Zeileis, and Annemarie A. Schuller. 2016. "Modeling Caries Experience: Advantages of the Use of the Hurdle Model." *Caries Research* 50(6): 517-26.

Javali, Shivalingappa B., and Parameshwar V. Pandit. 2010. "Using Zero Inflated Models to Analyze Dental Caries with Many Zeroes." *Indian Journal of Dental Research 21(4): 480-85.*

Khajedaluee, Mohammad et al. 2014. "The Burden of Diabetes in a Developing Country." Open *Journal of Preventive Medicine 04(04): 175-81.*

Noubiap J, Nansseu J , Endomba F, Ngouo A, Nkeck J, Nyaga U, Kaze A, Bigna J. 2019. "Active Smoking among People with Diabetes Mellitus or Hypertension in Africa_ a Systematic Review and Meta-Analysis." *Sci Rep* . 9(1): 588.

Ogbera, Anthonia Okeoghene. 2014. "Diabetes Mellitus in Nigeria: The Past, Present and Future." *World Journal of Diabetes* 5(6): 905.

Oguntibeju, O. O., N. Odunaiya, B. Oladipo, and E. J. Truter. 2012. "Health Behaviour and Quality of Life of Patients with Type 2 Diabetes Attending Selected Hospitals in South Western Nigeria." *West Indian Medical Journal* 61(6).

Olokoba, Abdulfatai B., Olusegun A. Obateru, and Lateefat B. Olokoba. 2015. "Type 2 Diabetes: A Review of Current Trends -." *International Journal of Current Research and Review* 7(18): 61-66.

Oritogun, Kolawole S, and Elijah A Bamgboye. 2018. "Application of Count Models on Infant and Child Mortality in Nigeria□ A Comparative Study." *International Journal of TROPICAL DISEASE & Health* 30(3): 1-12.

Oritogun KS, Oyewole O. O. and Odusan O. 2016. "Model Comparison and Factors Associated with Quality of Life of Type 2 Diabetic Patients: Gender Differentials." *British Journal of Medicine and Medical Research* 15(8): 1-12.

Oyewole, Olufemi O., Olatunde Odusan, and Ayotunde O. Ale. 2019. "Global Disability Burden and Its Predictors among Adult Nigerians Living with Type-2 Diabetes." *Ghana Medical Journal* 53(2): 135-41.

Peters, M. L., E. L. Huisman, M. Schoonen, and B. H.R. Wolffenbuttel. 2017. "The Current Total Economic Burden of Diabetes Mellitus in the Netherlands." *Netherlands Journal of Medicine* 75(7): 281-97.

Pittman, Brian et al. 2018. "Models for Analyzing Zero-Inflated and Over dispersed Count Data: An Application to Cigarette and Marijuana Use." *Nicotine & Tobacco Research:* 1390-98.

Ridoh, Diabetes Prevention and Control Program. 2010. "The Burden of Diabetes in Rhode Island." In Diabetes Prevention and Control Program·Www.Health.Ri.Gov, http://www.health.ri.gov/publications/burdendocuments/2010Diabetes.pdf.

Schram, Miranda, Caroline Baan, and Francois Pouwer. 2009. "Depression and Quality of Life in Patients with Diabetes: A Systematic Review from the European Depression in Diabetes (EDID) Research Consortium." *Current Diabetes Reviews* 5(2): 112-19.

Yusuf, Oyindamola B, Rotimi F Afolabi, and Ayoola S Agbaje. 2018. "Modelling Excess Zeros in Count Data with Application to Antenatal Care Utilisation." 7(3): 22-35.