

# Bayesian Models for Zero Truncated Count Data

---

## ABSTRACT

It is important to fit count data with suitable model(s), models such as Poisson Regression, Quasi Poisson, Negative Binomial, to mention but a few have been adopted by authors to fit zero truncated count data in the past, which were considered unsuitable. In recent times, dedicated models for fitting zero truncated count data have been developed by some authors which are improvements on the previous ones. Aside dedicated models for fitting count data; Bayesian multi-level was also recently developed to handle any type of count data. This study implemented Bayesian multi-level Poisson and Bayesian multi-level Geometric model, Bayesian Monte Carlo Markov Chain Generalized linear Mixed Models (MCMCglms) of zero truncated Poisson and MCMCglms Poisson regression model to fit health count data that is truncated at zero. Diagnostics tests were carried out and model selection criteria were used to determine preferred models for fitting the zero truncated data. The results showed that Bayesian multi-level Poisson outperformed Bayesian multi-level Poisson Geometric; also MCMCglms of zero truncated Poisson outperformed MCMCglms Poisson.

*Keywords: Count data, Bayesian inference, Health, insurance; Zero-truncated, Multi-level Models*

## 1. INTRODUCTION

Count data sets are type of data obtained by counting, and observations are taken within a fixed period of time; count data include zero and positive integers only. Example of studies where count data were modeled include the field of insurance [1], telecommunications [2], academics [3-4], medicine [5-9]. Other areas of studies include but not limited to agriculture, sports, biology and transportation.

Count data can be equi-dispersed, over-dispersed or under-dispersed. Over-dispersion is a problem when conditional variance is larger than the conditional mean, while under-dispersion is when the conditional variance is less than the conditional mean. Poisson model

is assumed to have equal mean and variance, therefore, it is not suitable to fit under-or over-dispersed count data [5]. Therefore, a more robust model is sought to achieve a more reliable estimates for data that exhibit over-or under-dispersion. Apart from observing the conditional mean and variance, another way to check for over-dispersion or under-dispersion is to carryout out a dispersion test, and this can be done by fitting quassi-Poisson regression model, [6] identified the robustness of quassi-poisson regression model in fitting count data. The feature of a given count data determines the type of model suitable for it; count data with many zeros can be modeled with zero-inflated or hurdle models, while count data with no zero count can be effectively modeled with zero truncated models. Studies such as [11-13] give details of such analysis.

Regression model for fitting count data have their background Generalized Linear Model (GLM), since in Linear model is considered inadequate in modelling count data because Linear model cannot account for heteroscedasticity in count data. GLM work in such a way that there is a link function that links the response variable to the predictors, and the count data need not to be normally distributed. The modes that belong to GLM belong to exponential class of family given in equation (1). A random variable  $Y$  has a distribution in the exponential family if its probability density function (pdf) has the form:

$$f(y|\theta) = c(y, \phi) \exp\{(y\theta - a(\theta))/\phi\}, \quad g(\mu_i) = x'\beta \quad (1)$$

Where 'y' is the value of an observation  $Y$ ,  $\theta$  is a location parameter called the canonical parameter,  $\phi$  is a dispersion parameter sometimes called the scale parameter and it determines the shape of the distribution,  $c(y, \phi)$  is a normalizing factor producing unit total mass for the distribution. Equation (1) was defined by [14], and the equation for  $f(y|\theta)$  indicates that the distribution of the response is in the exponential family.  $g(\mu_i) = x'\beta$  indicates that a transformation of the mean,  $g(\mu_i)$ , is linearly related to the explanatory variables contained in  $X$ . The data used for in this study was truncated at zero, so zero truncated Poisson and Geometric models are itemized in equations (2) to (4).

Let  $P_0(x; \theta)$  be the original Poisson distribution. Then the pdf of zero-truncated form of  $P_0(x; \theta)$  is given as follows

$$P_0(x; \theta) = \frac{P_0(x; \theta)}{1 - P_0(0; \theta)}; \quad x = 1, 2, \dots, n \quad (2)$$

And for zero-truncated binomial we have

$$P_N(x; \underline{\theta}) = \frac{1}{1-(1-p)^n} \binom{n}{k} p^k (1-p)^n \quad (3)$$

Where  $\underline{\theta}=(n; p)$ , and  $n \in \mathbb{N}, 0 < p < 1$ ,

While that of Zero truncated geometric distribution is given as

$$p(x; \theta) = \theta(1-\theta)^{x-1}, x = 1, 2, 3 \dots \quad 0 \leq \theta < 1 \quad (4)$$

The link functions for in fitting regression resulting from equation (2) and (4) is the log link, while that (3) is logit link.

This study proposed Bayesian multi-level model and MCMCglmm models to fit zero truncated count data, and identify the suitability of the models in fitting the data. The remaining part of this paper is sectioned as follows; section 2 contains Materials and Methods, section 3 is the Results and Discussion.

## 2. MATERIALS AND METHODS

### 2.1 Multi-Level Modelling

Multi-Level Modelling has to do with predicting the response variable ( $y$ ) using the linear combination  $\eta$  of predictors transformed by the inverse link function ( $f_i$ ). For a given distribution ' $d$ ', it can be written as

$$y_i \sim d(f_i(\eta_i), \theta) \quad (5)$$

where  $d$  is called family of the distribution as usually represented in statistical software, then the parameter  $\theta$  represents additional family parameters which naturally do not change as data points increases. We consider a general the linear predictor can be written as

$$\eta = \mathbf{A}\beta + \mathbf{B}\varepsilon \quad (6)$$

Where  $\beta$  and  $\varepsilon$  (fixed and random effect) are the coefficients at population-level and group-level respectively,  $\mathbf{A}$ ,  $\mathbf{B}$  are the corresponding design matrices. The response  $y$  as well as  $\mathbf{A}$  and  $\mathbf{B}$  makes up the data, while  $\beta$ ,  $\varepsilon$  and  $\theta$  are the model parameters to be estimated. Bayesian MCMC methods treat  $\varepsilon$  as a parameter unlike maximum likelihood which treats  $\varepsilon$  as error term [15]. Prior selection for Bayesian Multi-level modelling

corresponds to the use of No-U-Turn Sampler (NUTS) instead of Inverse-Wishart prior distribution as discussed by [15].

## 2.2 Generalized linear mixed models (GLMMs)

Generalized linear mixed models extension of Generalized Linear Models because as it incorporates fixed and random effects. The model can be simply put as follows:

$$y = X_i\beta + Z_i\gamma + e \quad (7)$$

Where  $y$  is a  $N \times 1$  column vector,  $X$  and  $Z$  are design matrices for the fixed and random predictors of the data respectively. These predictors have connected parameter vectors  $\beta$  and  $\gamma$ , and  $e$  is a vector of residuals. For random effects,  $\gamma_i = (\gamma_{i1}, \dots, \gamma_{iq})'$  explained the inclusion of covariates  $Z_i$ , the link can be written in extended form as

$$g(\mu_i) = \eta_i = X_i\beta + Z_i\gamma + e, \quad i = 1, 2, \dots, n \quad (8)$$

According to [1] MCMCglmm can be modelled using the  $R$ - and  $G$ -structure;  $R$  structure adopted in this study is random; and the latent variables are assumed to have the multivariate normal distribution and the conjugate prior of the variance structure is inverse-Wishart prior distribution which can be Gibbs sampled.

## 2.3 Parameter Estimation and Model Selection

It is important to carry out dispersion test in order to identify the type of dispersion a count data exhibits. Dispersion test is used to test the null hypothesis of equi-dispersion in Poisson GLMs against the alternative of true dispersion parameter of greater than 1. In standard Poisson GLM models, the (conditional) mean  $E(Y) = \mu$  is assumed to be equal to the variance  $Var(Y) = \mu$ . Dispersion test is used to assess the hypothesis that the assumption of equi-dispersion holds against the alternative that the variance is of the form:

$$Var(Y) = \mu + \alpha \times trans(\mu) \quad (9)$$

If  $trans=1$  or  $trans=2$ , it follows linear and quadratic formulation. Over-dispersion corresponds to  $\alpha > 1$  and under-dispersion to  $\alpha < 1$ . The transformation  $trans$  can either be specified as a function or an integer as defined by [16]. If  $trans$  is NULL, no formulation is used in the dispersion test and expressed as:

$$Var(Y) = (1 + \alpha)\mu = \mu \times (dispersion) \quad (10)$$

which translates to dispersion parameter for quasi-Poisson family.

After using quasi-Poisson regression analysis, further analysis was carried on the data with Bayesian multi-level analysis and MCMCglmm. As a default with Bayesian multi-level analysis, normal prior was adopted, using No-U-Turn Sampler (NUTS) to sample from the posterior distribution. 'brms' package by [17] the 'brms' package does no work in isolation but with stan processor. In order to determine the model with a better fit between Poisson and Geometric models in the context of Bayesian multi-level modeling, the Watanabe-Akaike Information Criteria (WAIC), [18] and Leave-one-out cross validation LOO-CV proposed by [19-20] were used, model with lower WAIC and LOO indicates a better fit for the data. On the hand, model selection criteria MCMCglmm are the Aikaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Deviance Information Criteria (DIC). The MCMC follows the Monte Carlo Part which relies on generation of pseudo-random numbers, and then the Markov chain part; a Markov chain is a sequence of number where each number is dependent on the previous number on the sequence, that is,

$$\theta_{t-1} \sim N(\theta_{t-2}, \sigma), \quad \theta_t \sim N(\theta_{t-1}, \sigma) \quad (11)$$

For example, considering a Gaussian proposal  $\theta_t = N(\mu, \sigma)$ , the proposal shifts to the right each time a sample is drawn from the distribution and plot generated from  $\theta_t$  is called trace plot, the density plot represents the sample generated. The trace and density plots of the predictors can be found under results and discussion.

For Bayesian Multi-level models, Pareto k analysis was carried out to determine if any observation was left out in the process of the analysis. Any observation with  $k > 0.7$  indicates a bad observation; consequently, the observation would be left out during analysis. Software package by [21] "package AER" by [16] was used to carry out the dispersion test. "trans"=NULL was adopted in this study.

#### **2.4 Data Description**

The dataset used in this article consist of National Health Insurance Scheme (NHIS) data with no zero count. The data was obtained from health facility in Ota, Ogun State, Nigeria. A sample consisting of 1647 users of NHIS was obtained from July 2016 to July 2017. The response variable is "Number of Encounter (Nencounter)", while predictor variables are Sex, Age of patients, Number of diagnosis (Ndiagnosis) for the period of visits, individual on follow-up (Follow-up) and Eclass (In-patient or out-patient). The data is under-dispersed with dispersion parameter of 0.7806 (the dispersion parameter is less than 1). The data is further described as follows:

Response variable (Nencounter) that is, number of times claims were made. The class (Eclass) indicated whether a patient was ever or not on admission for the period, that is, (in-patient=1, out-patient= 0). Another predictor is (follow-up), indicating whether a patient is on regular check-up or not, (follow-up=1, no follow-up=0). Gender (sex) of patients; (male=1, female=0). Another predictor is Ndiagnosis, which represents the number of diagnosis a patient had for the period of encounter. The last predictor included is biological age of patient.

The classification of gender according to the number of times patients had Encounter(s) at the health facility is presented in Table 1.

**Table 1:** Classification of gender according to Number of Encounter

Number of Encounter	Sex		Total
	Female	Male	
1-5	745	630	1375
6-10	105	91	196
11-15	31	21	52
16-21	6	9	15
21-27	7	2	9
Total	894	753	1647

Source: Authors' Computation

From Table 1, there is an indication that females had more encounter as compared to their male counterpart. The information on patients that are either on follow up or not is represented in Table 2.

**Table 2:** Information on follow-up status

NEncounter	Follow-up		Total
	Female	Male	
1-5	1148	227	1375
6-10	163	33	196
11-15	41	11	52
16-21	12	3	15
21-27	6	3	9
	894	753	1647

Source: Authors' Computation

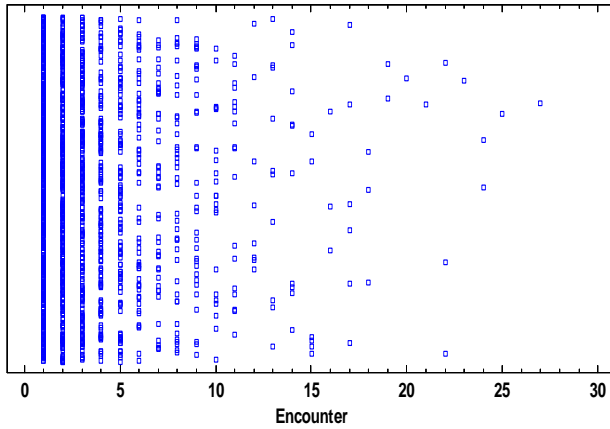


Figure 1: Scatter plot of Nencounter

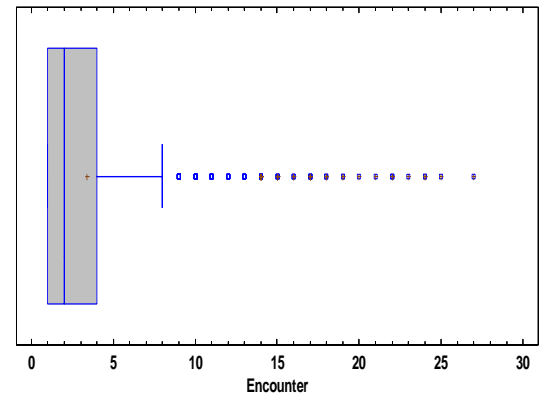


Figure 2: Box and Whisker plot of Encounter

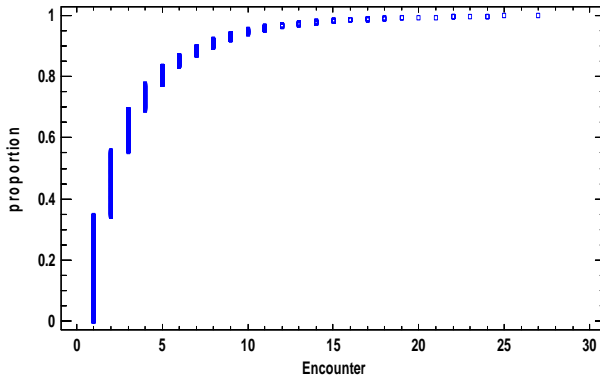


Figure 3: Quantile plot of encounter

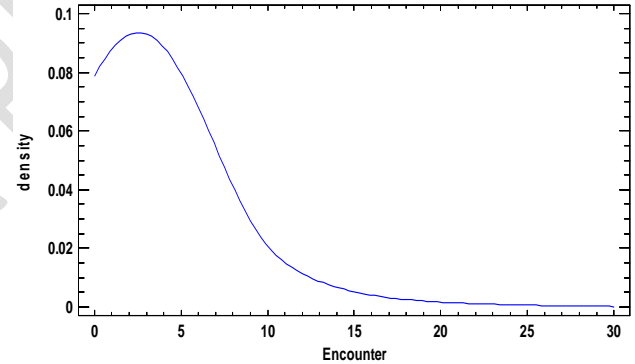


Figure 4: Density trace plot of Nencounter

### 3. RESULTS AND DISCUSSION

The result for the Quasi-Poisson regression analysis is presented in Table 3.

Table 3: Quasi-Poisson Regression Analysis

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.2941138	0.0295171	9.964	<2e-16 ***
Sex	0.0111539	0.0239250	0.466	0.64113
Age	0.0018506	0.0006451	2.869	0.00417 **
followup	-0.1534049	0.0324182	-4.732	2.41e-06 ***
Eclass	0.1700727	0.0955086	1.781	0.07515 .
Ndiagnosis	0.2677050	0.0037693	71.023	<2e-16 ***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Source: Authors' Computation

From the dispersion test carried out using quasi Poisson regression, dispersion parameter is 0.7806, indicating that the data is under-dispersed as shown in Table 3,  $z = -3.0491$ , ( $P=0.9989$ ), the dataset is under-dispersed since  $\alpha < 1$ .

Table 4 shows that only follow-up have significantly negative effect on Encounter, so whether a patient is on follow-up or not does not necessarily mean that encounter will increase. With quasi-Poisson, [6] identified that the inadequacy inherent in Poisson model is taken care of. Model selection for Bayesian multi-level is performed in Table 4, and \* indicates model with lower value between Bayesian multi-level Geometric and Bayesian multi-level Poisson.

Table 4: Bayesian Multi-level Model Selection

Model	elpd_waic	p_waic	waic	elpd_loo	p_loo	looic	Waic=LOO
Geometric							
Est.	-3534	0.9	7069.1	-3534	0.9	7069.1	Yes
SE	30	0.1	60.1	30	0.1	60.1	
Poisson							
Est.	-2862.0	18.2	5724.4*	-2862.0	18.6	5724.8*	No
SE	56.3	7.2	112.5	56.4	7.4	112.8	

Source: Authors' Computation

All Pareto k estimates are good ( $k < 0.5$ ) in the case of Geometric, but for Poisson model, two observations are bad with estimate of  $k > 0.7$ .

Table 4 shows that Bayesian Multi-level model with Poisson outperformed geometric based on the WAIC and LOO estimates, contrary to maximum likelihood estimates for Poisson. The results for population-level effects model for Geometric and Poisson models are presented in Tables 5 and 6.

Table 5: Population-Level Effects model for Bayesian Multi-level Geometric:

	Est.	Error	l-95% CI	u-95% CI	Eff. Sample	Rhat
Int.	-.03	.08	-.18	.12	4244	1.00
Eclass	.05	.25	-.41	.54	3908	1.00
Followup	.01	.08	-.14	.16	4260	1.00
Sex	.00	.06	-.11	.11	3868	1.00
Age	.00	.00	-.00	.00	5059	1.00
Ndiagnosis	.37	.02	.34	.40	4109	1.00

Source: Authors' Computation



Table 6: Population-Level Effects model for Bayesian Multi-level Poisson

	Est.	Error	l-95% CI	u-95% CI	Eff.Samp	Rhat
Intercept	.29	.03	.22	.36	3624	1.00
Eclass	.16	.11	-.05	.37	2606	1.00
followup	-.15	.04	-.23	-.08	3283	1.00
Sex	.01	.03	-.04	.07	3159	1.00
Age	.00	.00	.00	.00	4203	1.00
Ndiagnosis	.27	.00	.26	.28	3586	1.00

Source: Authors' Computation

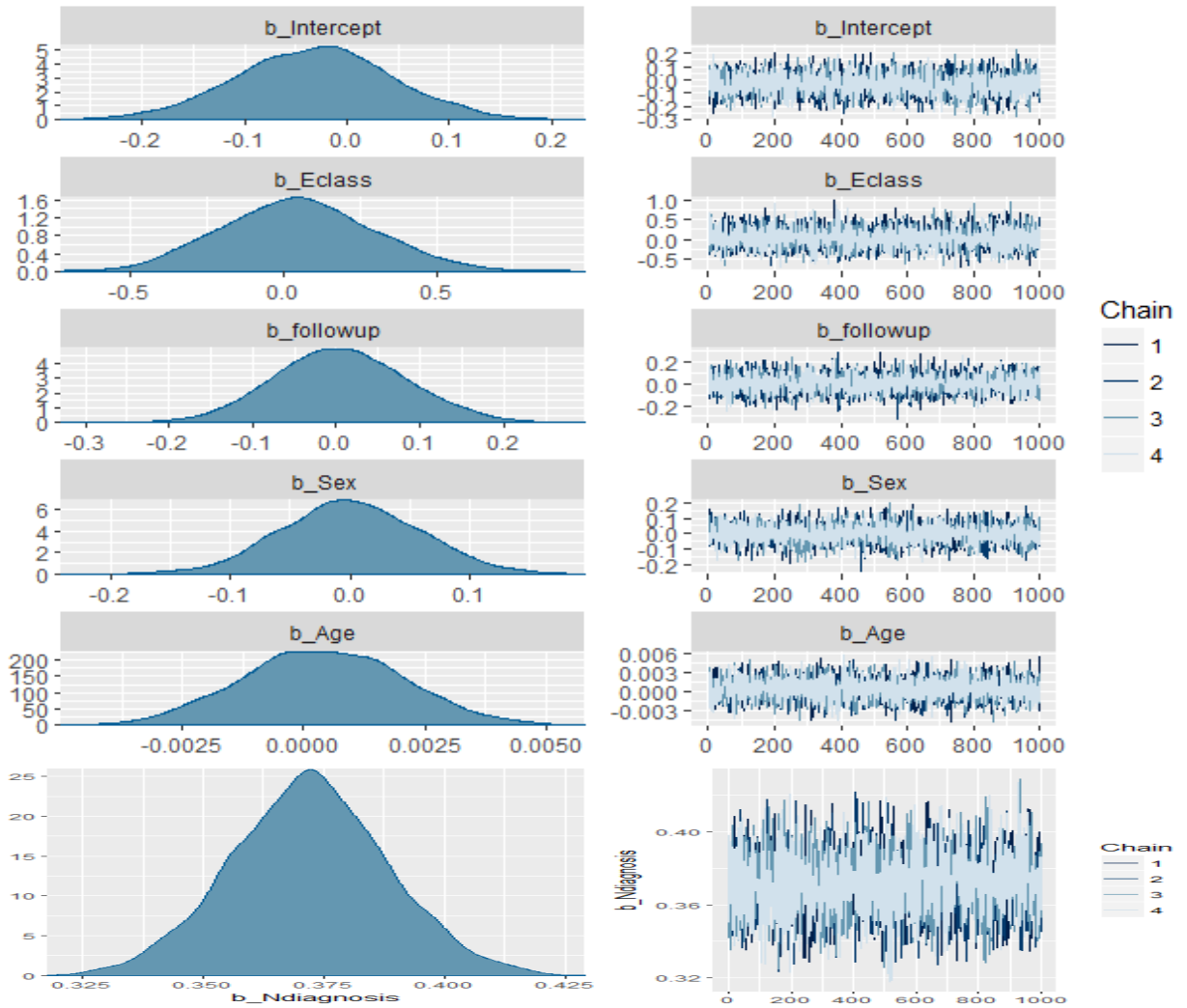
Table 7: Bayesian MCMCglmm Zero truncated Poisson and Ordinary Poisson regression Model

Model selection	Bayesian ZT Poisson	Bayesian Poisson
AIC	4709.586	5463.467
BIC	4747.433	5501.314
DIC	5027.495	5630.394

Source: Authors' Computation

Samples were drawn using sampling (NUTS). Scale reduction factor on split chains (at convergence, Rhat = 1). Log link was used for both Bayesian Geometric and Poisson models, the 'Est.' in column 2 of Table 5 and Table 6 represents the posterior mean, while 'Error' stands for standard deviation of the posterior mean, 'CI' stands for Confidence Interval, and 'Eff.Samp' stands for Efficient Sampling. All the parameters have positive relationship with "NEncounter" showing that number of Encounter can be determined by each of the predictor. By implication, patients on admission (inpatients) have more encounter than outpatients, since it was coded (0, 1), patients on follow-up made more encounter than patients that are not on follow-up, since it was coded (0, 1). Also, male patients made more encounter than female, coded as (male=0, Female=1), increase in age leads to increase in number of encounter. From Table 5, Ndiagnosis (0.37) account for number of Encounter than Eclass, Follow-up, Sex, and Age.

From Table 6 for Bayesian multi-level Poisson model, All the parameters have positive relationship with Encounter, except 'follow-up', showing that number of Encounter can be determined by each of the predictor but not in the case of follow-up; "follow-up" has significantly negative effect on "Encounter" as identified in Table 3 in the case of classical Quassi Poisson regression analysis, therefore, whether a patient is on follow-up or not does not necessarily mean that Encounter will increase. The plots for the results are presented in Figure 5.



**Figure 5:** Trace and Density plots of all relevant parameters for Bayesian Multi-level with Geometric model

**Table 8:** Posterior Mean for Bayesian MCMC glmm for Zero Truncated Poisson Model

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	-0.2365188	-0.3428029	-0.1292316	678.7	<3e-04 ***
Sex	0.0042738	-0.0675917	0.0813161	1020.5	0.9137
Age	0.0014956	-0.0004543	0.0037423	1281.2	0.1545
Ndiagnosis	0.3638062	0.3455116	0.3798459	638.7	<3e-04 ***
followup	-0.0947547	-0.1988362	0.0025938	1159.1	0.0669
Eclass	0.1145752	-0.1929133	0.4219337	1283.6	0.4341

Source: Authors' Computation

In this study Bayesian multi-level and Bayesian MCMCglms have been employed to fit zero truncated count datasets from health and insurance domain, basic descriptive analysis was carried out on the data and quasi-Poisson regression analysis was equally performed, first to determine the dispersion type and second, to determine the relationship of the predictors with the response count variable. The quasi-Poisson analysis shows that it is only follow-up that did not necessarily have relationship with Encounter among all the variables. Bayesian Multi-level regression analysis was implemented and the result shows that Bayesian Multi-level Poisson regression outperformed Geometric model using “waic” and “looic” as presented in Table 4, and in the class of Bayesian MCMCglms, Bayesian Zero truncated Poisson model outperformed Ordinary Poisson regression Model using AIC, BIC and DIC as presented in Table 7.

For the Bayesian multi-level model based Geometric distribution, all the predictors have positive relationships with “Encounter”, while Bayesian multi-level model based Poisson distribution showed the same relationship with that of quasi-Poisson model, indicating that of follow-up is negative. That further shows how reliable quasi-Poisson is in fitting count datasets. The posterior mean (estimate) for MCMCglmm Zero truncated Poisson model (-0.0947547) and that of ordinary Poisson regression Model (-0.0903413) using R-structure, as detailed in [3], also have all estimates to be positive except that of follow-up, as shown in Table 8. The result containing that of MCMCglms Poisson was not tabulated but reported. The result agreed with that of the Multi-level Poisson model. These results also agreed with [3] and [22] where authors demonstrated that Bayesian models based on Generalized Linear Mixed Models (MCMCglms) are suitable to fitting count data effectively, although [2] demonstrated the superiority of Dirichlet Prior Mixture Model over MCMCglms.

Other deductions from the study shows that more females are on hospital admission than their male counterpart, also patients on follow-up are far more than patients not on follow-up, and this might indicate that more people are aware of the need to follow-up on their treatments and health status.

From the results obtained, it can be deduced that when comparing Bayesian Multi-level Poisson with Geometric model, Poisson outperformed the Geometric. On the other hand, when comparing Bayesian zero truncated family of MCMCglms with MCMCglms Poisson, the zero truncated family outperformed the Poisson family. The four models cannot be compared at the same time because the model selection criteria for Bayesian Multi-level models are WAIC and LOO ([18-20]), which is considered an improvement on DIC.

Therefore, when a researcher is considering a suitable multi-level model to fit under-dispersed count data that is truncated at zero, Poisson model should be considered, the problem associated with classical Poisson model has been taken care of when using Bayesian multi-level models because Bayesian multi-level models have unique sampler and unique prior selection technique ([23-24]), which is considered robust for any type of count data.

## References

- [1] Jong P, Heller GZ. Generalized Linear Models For Insurance Data, 1<sup>st</sup> ed., Cambridge University Press, New York, ISBN-13 978-0-511-38677-0, 2008.
- [2] Paleti R. Generalized Extreme Value models for count data: Application to worker telecommuting frequency choices. *Transportation Research Part B: Methodological*, 2016, 83: 104-120.
- [3] Adesina OS, Agunbiade DA, Osundina SA. Bayesian Regression Model for Counts in Scholarship. *Journal of Mathematical Theory and Modelling*, 2017;7(9): 46-57
- [4] Adesina OS, Olatayo TO, Agboola OO, Oguntunde PE. Bayesian Dirichlet Process Mixture Prior For Count Data . *International Journal of Mechanical Engineering and Technology*, 2018; 9(12): 630-646
- [5] Cameron AC, Johansson P. *Count data regression using series expansion: With applications*, J. Appl. Econom, 1997; 12:203–223.
- [6] Cameron AC, Trivedi PK. Count Panel Data. *Oxford Handbook of Panel Data Econometrics*, Oxford University Press, 2013.
- [7] Kokonendji CC. Over- and under-dispersion models, in *Methods and Applications of Statistics in Clinical Trials, Volume 2: Planning, Analysis, and Inferential Methods*, N. Balakrishnan, ed., Wiley, Hoboken, NJ, 2014: 506–526.
- [8] Haselimashhadi H, Vinciotti V, Yu K. A new Bayesian regression model for counts in medicine. *arXiv:1601.02820 [stat.ME]*; 2016:1-21
- [9] Workie MS, Lakew AM. Bayesian count regression analysis for determinants of antenatal care service visits among pregnant women in Amhara regional state, Ethiopia. *Journal of Big Data*, 2018; 5: 7. <https://doi.org/10.1186/s40537-018-0117-8> 5:7.
- [10] Cameron AC, Trivedi PK. Regression Analysis for Count Data, *Econometric Society Monograph*, 2<sup>nd</sup> ed., Cambridge, UK, Cambridge University Press, 1998.
- [11] Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing, *Technometrics*, 1992; 34: 1–14.
- [12] Famoye F, Wang W. Censored generalized Poisson regression model, *J. Comput. Stat. Data Anal.*, 2004; 46:547–560.

- [13] Famoye F, Singh KP. Zero-inflated generalized Poisson regression model with applications to domestic violence data, *J. Data Sci.*, 2006; 4:117–130.
- [14] Nelder JA, Wedderburn RWA. Generalized Linear Models, *Journal of the Royal Statistical Society, Series A (General)*, 1972; 135(3):370–384.
- [15] Fox J, Weisberg S. An R companion to Applied Regression, Second Edition. Sage, 2011.
- [16] Kleiber C, Zeileis A. Applied Econometrics with R. New York: Springer-Verlag, ISBN 978-0-387-77316-2, 2008 URL <https://CRAN.R-project.org/package=AER>
- [17] Burkner PC. brms: An R Package for Bayesian Multilevel Models using Stan." *Journal of Statistical Software*, 2017.
- [18] Watanabe S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *The Journal of Machine Learning Research*, 2010; 11:3571-3594.
- [19] Gelfand AE, Dey DK, Chang H. Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods." Technical report, DTIC Document, 1992.
- [20] Vehtari A, Gelman A, Gabry J. Efficient Implementation of Leave-One-Out Cross-Validation and WAIC for Evaluating Fitted Bayesian Models." Unpublished manuscript, 2015; 1-22. URL [http://www.stat.columbia.edu/~gelman/research/unpublished/loo\\_stan.pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/loo_stan.pdf).
- [21] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2018, <https://www.R-project.org>
- [22] Brooks ME, Kristensen K, Benthem KJ, A Magnusson, CW Berg, A Nielsen, HJ Skaug, M Maechle, M Bolker. Modelling Zero-Inflated Count Data With glmmTMB. bioRxiv preprint bioRxiv:132753; doi:<https://doi.org/10.1101/132753>; 2017.
- [23] Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *The Journal of Machine Learning Research*, 2014; 15(1):1593-1623.
- [24] Lewandowski D, Kurowicka D, Joe H. Generating Random Correlation Matrices Based on Vines and Extended Onion Method." *Journal of Multivariate Analysis*, 2009; 100(9):1989-2001.