# Statistical models for longitudinal zero-inflated count data: application to seizure attacks

Fenta Haile Mekonnen[1], Workie Demeke Lakew[1], Zike Dereje Tesfaye[1], Prafulla Kumar Swain[2]

1. Department of Statistics, College of Science, Bahir Dar University, Bahir Dar- Ethiopia.
2. Department of Statistics, Utkal University, Bhubaneswar-751004, India.

**Emails:** hailemekonnen@gmail.com; demay_gu06@yahoo.com; derejetesfaye11@gmail.com; prafulla86@gmail.com

## Abstract

**Background:** Chronic non-communicable diseases:- such as epilepsy, are increasingly recognized as public health problems in developing and African countries. This study aimed at finding determinants of the number of epileptic seizure attacks using different count data modeling techniques.

**Methods:** Four common fixed-effects Poisson family models were reviewed to analyze the count data with a high proportion of zeros in longitudinal outcome, i.e., the number of seizure attacks in epilepsy patients. This is because, in addition to the problem of extra zeros, the correlation between measurements upon the same patient at different occasions needs to be taken into consideration.

**Results:** The investigation remarkably identified some important factors associated with epileptic seizure attacks. As people grow old , the number of seizure attacks increased and male patients had more seizures than their female counterparts. In general, a patient's age, sex, monthly income, family history of epilepsy andservice satisfaction were some of the significant factors responsible for the frequency of seizure attacks (P value<0.05).

**Conclusion:** This study suggests that zero-inflated negative binomial is the best model for predicting and describing the number of seizure attacks as well as identifying the potential risk factors. Addressing these risk factors will definitely contain the progression of seizure attack.

**Keywords:** linear mixed model, hurdle model, seizure attacks, zero-inflated models.

## Background

Chronic non-communicable diseases, such as epilepsy, are increasingly recognized as public health problems in developing countries. Even if, there is a cheap and effective treatment, the majority of people with epilepsy remain untreated. Epilepsy is one of the most common and serious brain disorders affecting at least 50 million people worldwide. People with epilepsy have a mortality rate 2-3 times higher than the general population[1]. Educational achievement, employment rate, and quality of life are all substantially lower for people with epilepsy[2]. In spite of recentglobal advancements in diagnosis and treatment, about eight million people with epilepsy in Africa are not treated with modern antiepileptic drugs as a result of whichhas to be seen from a public health perspective[3,4].

Five controlled incidence studies have been conducted in African countries. These studies showedthat the number of new cases of epilepsy detected among 100,000 people during one year were: 83 in Burkina Faso, 64 in Ethiopia, 73 in Tanzania, 119 in Togo, and 156 in Uganda. These incidence rates are higher than those reported from the developed world, which usually range from 40 to 70 per 100,000people. When interviewed neurologists in the African region, usually report that epilepsy is the second or third reason for consultation after headaches or pe-

**Corresponding author:**
Fenta Haile Mekonnen,
Department of Statistics,
College of Science,
Bahir Dar University, Bahir Dar- Ethiopia.
Email: hailemekonnen@gmail.com

ripheral neuropathies, and the second or third reason for hospitalization after strokes and/or spinal cord pathologies[5]. Additionally, the prevalence and incidence rates of epilepsy in rural and suburban areas are usually higher than in cities. This is due to the fact that risk factors are more concentrated in the rural and the suburban areas where migrants from villages to cities reside in very poor conditions.[6]

The prevalence of epilepsy is substantially increasing in Ethiopia, including the specific study area. As far as our knowledge is concerned, there is no published paper that documented the determinant factors affecting the progression and change of the number of epileptic seizure patients. Therefore:- this study addressed the objectives by applying longitudinal data analysis, especially count data regression modelling, to identify the potential factors associated with epileptic seizure attacks. Someprevious literatures provide a plausible analysis of various problems. For example, analyzed longitudinal data of epileptic seizure counts using hidden Markov regression approach, and[8] demonstrated the use of longitudinal count data modelling in childhood injuries. Over the past years, different authors have suggested several predictors of seizure episodes such as age, gender, family history of epilepsy and seizure characteristics[9]. However, most of the scholars considered chi-square and logistic regression analysis[10,11]. Epileptic seizures have a historical association with religion and sprit possession. some scholars stated that, the development of epilepsy, is related with socio-economic factors such that infection, injury, poor nutrition, low educational achievement and poor housing status[12,13]. About 36% of Africans live in cities. Most of these city dwellers live in the suburbs in poor conditions characterized by over-crowding, poor water supply and bad sanitation. Consequently, there is a high prevalence of communicable diseases such as malaria, meningitis, cysticercoids and tuberculosis:- all of which are the frequent causes of epilepsy.

## Statistical analysis of count outcomes data

In medical research, count data are common. In addition, these event count data are often positively skewed with a high proportion of zeros. In these case, traditional methods may not always be appropriate for count data with extra zeros. The zero adjusting models have been used to analyze count data with extra-variability and excessive zeros. This model can be zero-inflated or hurdle models[14-17]. More recently, these modeling techniques have been extended to longitudinal and multilevel settings to accommodate correlated and incomplete count data[18,19].

Class of models for analyzing count data:- An overview: Inthetheoretical background of probability models, count data usually have Poisson distribution. Several modifications (extensions) to the Poisson model have been proposed for different reasons: (1) repeated measures of the outcome variable, (2) the occurrence of over-dispersion (the variability of the data is larger than the mean), and (3) the occurrence of excess zeros. The first and the second reasonscan be managed respectively, by including the random subject specific-effects[20] and by the over-dispersion model, like negative binomial for count data[21]:- where the natural parameter is assumed to follow a gamma distribution. Moreover, an excessive number zero, the third reason, can be accounted by including zero-inflated models[14]. Therefore for these reasons, four modified Poisson models: the negative binomial (NB) model, the zero-inflated Poisson model (ZIP), the zero-inflated negative binomial model, and the hurdle Poisson model[22] were introduced.

Poisson                      NB

*Accessing zero*                                      *Existing zeros + over-dispersion*

ZIP                          ZINB

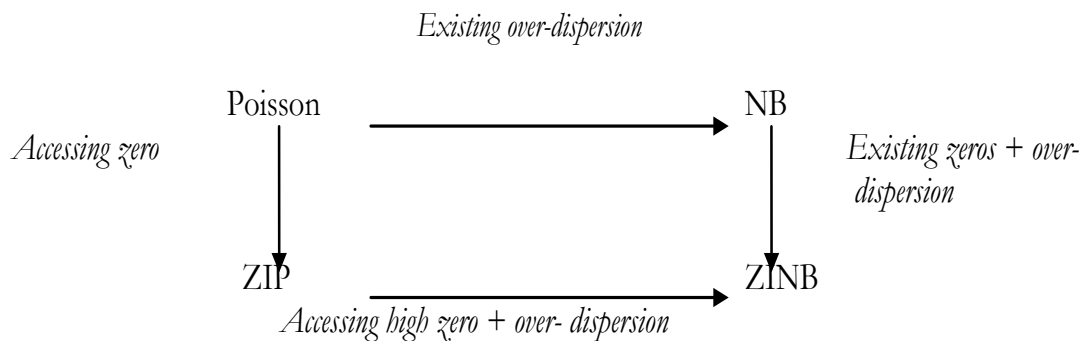*Accessing high zero + over- dispersion*

**Fig. 1:** Frequently used models in the count data analysis framework.

The aforementioned count models are fixed-effects models for independent (cross-sectional) count data. These models have been extended to longitudinal data. Longitudinal data are often complicated by at least three factors: the within-subject observations are usually not independent due to repeated measures on the same subjects (cluster); the between-subject variation may not be constant overtime; and the data are often incomplete, as subjects may drop out at any follow-up time.

Currently, there are two main approaches for modeling longitudinal and other correlated data, including count data. These are generalized estimating equations (GEE),[23] and generalized linear mixed-effects models(GLMM)[24]. The key differences between the GEE and GLMM are the following:-firstly, GEE is a marginal orpopulation average model that centers on correct estimation of the regression parameters by treating the within-subject (cluster) correlation as a "nuisance" with its robust estimation feature. In contrast, GLMM is a subject/cluster-specific model in which the subject/cluster-specific effects and the variance-covariance matrix are essential parts of the model. Secondly, GEE has a more restrictive assumption about missing values (i.e., missing completely at random, CAR) than that of the GLMM model (i.e., missing at random, or MAR)[25]. Because GLMM or the generalized multilevel models are more popular among social scientists and applied researchers, and because its assumption about the missing value mechanism is more realistic, this paper focusedon the GLMM approach for modeling the number of seizure attacks data using the aforementioned five modeling strategies. The previous models for cross-sectional count data discussed earlier can be viewed as part of the generalized linear model (GLM)26 for modeling cross-sectional data. GLM extends ordinary least squares regression by allowing the outcome probability distribution to be any member of an exponential family of distributions for count data. Generalized linear mixed-effects models (GLMM) extend GLM to model correlated (longitudinal) data and incomplete data27. The GLMM approach accounts for the within-subject correlations due to clustering by eitherincluding random effects into the model or by directly modeling these correlations with the appropriate variance–covariance structure. The linear mixed-effects/multilevel/hierarchical model28 is a special case of GLMM where the response is a continuous variable. A longitudinal count outcome, depending on its distribution, is analyzed with either the Poisson model, the negative binomial (NB) model, or one of the zero-adjusting models in the GLMM framework.

Under the longitudinal setting and assuming independence between subjects, let the random variable Yij denote the longitudinal response (such as the seizure attacks) for subject i=1, 2,..n, at the time j, where j=1, 2,… and it is assumed to follow the Poisson distribution with mean ij. The marginal mean is regressed on a set of covariates Xi=(xi1, xi2,…xip)' using a log-link.

$$y_{ij} \sim Poisson(\lambda_{ij}) \ and \ \log(\lambda_{ij}) = X_i'\beta \ ...(1)$$

where β is a vector of parameter associated with the vector covariates Xi. The Poisson regression model is assumed in its simplest form, the marginal mean and variance of the response are equal. This strong assumption, often not tenable for empirical data due to heterogeneity introduced in the data when important covariates are omitted from the study, is relaxed by applying an over-dispersed model. A commonly used over-dispersed model

is the negative binomial model. The negative binomial model is expressed as:

$$f(Y_{ij} = y_{ij}) = \frac{\overline{)(a + y_{ij})}}{\overline{)(a)}y_{ij}!}\left(\frac{b}{1+b}\right)^{y_{ij}}\left(\frac{1}{1+b}\right)^{a}\ ...(2)$$

In this model, a-1 is called the over-dispersed parameter due to unobserved heterogeneity and λij is the mean number of seizure attacks. The negative binomial regression model can be obtained similar to equations (1) by using equation (2).

The NB model may not be appropriate if the over-dispersion is due to excess zeros because it underestimates the probability of zeros and consequently underestimates the variability present in the outcome. In such situations, alternative models such as zero-inflated/hurdle models that account for over-dispersion due to excess zeros are useful.

The zero-inflated Poisson (ZIP)24 has been proposed to address this issue. The model assumes that counts are rather generated by two processes. The first process generates zero counts with probability пij while the non-zero counts follows the Poisson distribution with parameter λij and are realized with probability (1- пij). In addition, the model assumes that zero counts are generated from two sources based on the probabilities of the two processes. Thus, the probability distribution function of the longitudinal ZIP model can be written as:

$$Y_{ij} \sim \begin{cases} 0 & with\ probability\ \Pi_{ij} + (1 - \Pi_{ij})\exp^{-\lambda_{ij}} \\ y_{ij} & with\ probability\ (1 - \Pi_{ij})\dfrac{\lambda_{ij}^{y_{ij}}\exp^{-\lambda_{ij}}}{y_{ij}!}\ ...(3) \end{cases}$$

This model is reduced to the Poisson model when пij=0. Note that within this model, all zero counts are typically described though logistic regression, whereas positive counts are realized by log-linear Poisson model. For the vector of covariates Zij and Xij with associated parameters α and β respectively, the model specifications are expressed as follows:

$$Logit(\Pi_{ij}) = \alpha' Z_{ij}\ and\ \log(\lambda_{ij}) = \beta' X_{ij}\ ...(4)$$

The ZIP model accounts for over-dispersion due to excess zeros, but not due to unobserved heterogeneity. Moreover, if the count process does not follow the Poisson model then one may use zero-inflated negative binomial (ZINB) model29, by considering the count process as negative binomial distribution. In contrast to ZIP model, the ZINB model is more appealing since an additional parameter captures variability due to over-dispersion. Thus, the probability distribution function of ZINB is given by:

$$Y_{ij} \sim \begin{cases} 0 & with\ probability & \Pi_{ij} + (1 - \Pi_{ij})\left(\dfrac{a^{-1}}{a^{-1} + \lambda_{ij}}\right)^{a^{-1}} \\ y_{ij} & with\ probability & (1 - \Pi_{ij})\left[\dfrac{\overline{)(y_{ij} + a^{-1})}}{\overline{)(a^{-1})}y_{ij}!}\left(\dfrac{a^{-1}}{a^{-1} + \lambda_{ij}}\right)^{a^{-1}}\left(\dfrac{\lambda_{ij}}{a^{-1} + \lambda_{ij}}\right)^{y_{ij}}\right]...(5) \end{cases}$$

Here the parameters пij and λij have a similar meaning as in ZIP model. The zero-inflated negative binomial regression model can be obtained similar to equations (4) by using equation (5).

## Methods
### Data collection
Doctors and BSc psychiatry nurses who worked in psychiatric clinic at Felege Hiwot Referral Hospital were selected and trained to collect data. In the hospital, epileptic patients get follow-up from psychiatrynursesand the epilepsy clinic every month. The hospital is located in Bahir Dar city, Ethiopia, 565 km from Addis Ababa, the capital of Ethiopia. A structured questionnaire was used to collect data on social, demographic, behavioral and economic factors. To ensure the quality of data, the researchers were engaged in continuous supervision and monitoring.

### Model variables and measurements
A total of n=53 patients' demographic details were included for this analysis. The outcome was measured using different measurements fromeach patient's card, and the covariates were selected from the various aspects of the collected data. The covariates were classified as time-independent and time-varying.

### Time-independent covariates
Baseline demographic and socio-economic variables included in this analysis are age (in years); marital status (0=never married, 1= ever married); sex (0= female, 1 = male); place of residence (0 = urban, 1 = rural); presence of other diseases (0=no, 1=yes); family history of epilepsy (0=no, 1=yes);monthly income; diagnostic period length (in years); service satisfaction; family support (0=no, 1=yes); employment status; age of on set (age at which the disease started); history of any trauma (accident) (0=no, 1=yes) and types of seizures.

### Time-varying covariates
Time was be measured as a continuous variable representing monthly follow-up visits to the hospital. The variable time starts with the value 0 for the first follow-up visit, 1 for the second visit, and so on.

## The dependent variable

The number of seizure attacks per month (seizure attacks per month)

## Model evaluation and selection

It is known that, the Poisson distribution is nested within negative binomial (NB) when the dispersion parameter is zero, and it is nested within zero-inflated Poisson (ZIP) when there is no zero-inflation. Moreover, NB is nested within ZINB when there is no zero-inflation and ZIP is nested within ZINB when the dispersion parameter is zero. Thus, for nested models, the differences in -2LogLikelihood (-2LL) between each set of nested pairs can be used to evaluate the best model fit. In addition, a more convenient way to compare the model fit across all models is to use information criteria: Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC)[30,31]. The p-values less than 5% were considered as significant results. STATA 12.0 package was used for all statistical analyses.

## Results

## Descriptive and exploratory analyses

The outcome variable: the number of seizure attacks per follow-up periods ranged from 0 to 30. The average number of attacks in each of the follow-up periods ranges from 0.3 to 6.03, and it was substantially smaller than the variance (0.46 to 66.59), which indicated that there exists over-dispersion in the data set. Half of the patients (50%) had at least 3 seizure attacks before starting the medication and three-fourth (75%) of them had at most 2 seizures at the third visit (Table 1). Taking the data set all together, the variance (17.19) was 11.6 times higher than the mean (1.49), which is a clear implication of the existence of over-dispersion. Moreover, the distribution of the data was highly positively skewed (skewness=5.65), with a high spike on the left and a long tail on the right. Values ranged from 0 to 30, but about 94% of the counts were between 0 and 4. Generally, there were about 61.5% zeros (Fig. 1), indicating excess zeros in the data.

**Table-1:** Descriptive Statistics for outcome variable
(Number of seizure attacks per follow-up)

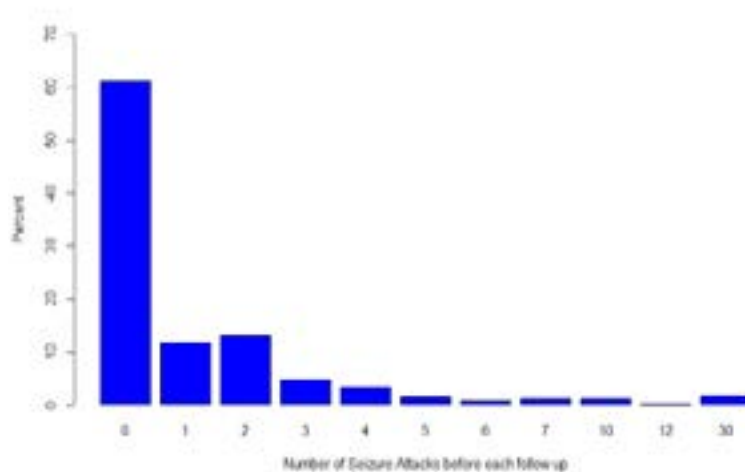| Follow-up periods | Mean | Variance | Minimum | Maximum |
|---|---|---|---|---|
| Baseline | 6.03 | 66.59 | 1 | 30 |
| 1 | 3.12 | 52.13 | 0 | 30 |
| 2 | 1.18 | 66.59 | 0 | 7 |
| 3 | 0.82 | 3.53 | 0 | 5 |
| 4 | 0.79 | 1.54 | 0 | 4 |
| 5 | 0.42 | 1.61 | 0 | 3 |
| 6 | 0.3 | 0.46 | 0 | 2 |
| 7 | 0.3 | 0.90 | 0 | 4 |
| 8 | 0.3 | 1.04 | 0 | 4 |

**Fig. 1:** Frequency distribution of the number of seizure attacks

All four models (Mixed effects Poisson, NB, ZIP and ZINB) fitted to the data. The results of all the models are summarized in Table 2. To select the best model fit for the data, we used formal evaluation methods, including: the likelihood value statistic ($-2LL$), the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC),which are used to measure how far off the fitted model is from the observed data sets. Thus, the smaller value indicated a better fit. The dispersion parameters for NB and ZINB were 0.86 and 1.8 respectively which were highly significant (p-value<0.01), implying, in their order, that the NB is superior to Poisson and ZINB is again superior to the ZIP. Comparing the four models together, the -2LL and the two information criteria (AIC and BIC) listed at the bottom indicated that the Poison model had the poorest fit, and zero-inflated negative binomial had the best fit (-LL=. 296, AIC=702.1and BIC=776.5).Since the zero-inflated negative binomial (ZINB) model had

the best fit, all the interpretations were from this model. Contrary to the expectations, marital status, educational background, employment status, place of residence and other factors which are listed in the table were not associated with the frequency of the number of seizure attacks of the epileptic patients. However, age, sex (being male) and having a family history of epilepsy were positively and significantly associated with the number of seizure attacks. Specifically, conditional on random effects, when the age of the patient increased by one year, the frequency of seizure attacks on average will be exp(0.07)=1.07, holding all other factors (variables) constant. The significant negative time trend (exp(-0.79)=0.45) showed that, as the follow-up times increased regularly, the number of seizure attacks decreased almost by half. The significant random effects (1.8 with p-value<0.05) revealed considerable heterogeneity among individuals with respect to the frequency of seizure attacks (Table 2).

**Table 2:** Results of mixed-effects Poisson, negative binomial (NB), zero-inflated poison (ZIP), zero-inflated negative binomial (ZINB), and hurdle models.

| Variables | Poisson | | NB | | ZIP | | ZINB | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | S.E. | Estimate | S.E. | Estimate | S.E. | **Estimate** | **S.E.** |
| Age | 0.062 | 0.026*** | 0.063 | 0.02** | -0.17 | 0.09** | **0.07** | **0.06**** |
| Sex (male) | 2.84 | 0.44*** | 2.56 | 0.61*** | -5.7 | 2.5* | **2.5** | **0.92**** |
| Marital status(Married) | -2.48 | 0.67*** | -2.56 | 1.03* | 7.31 | 3.90 | **-2.20** | **1.79** |
| Education (literate) | 0.37 | 0.40 | 0.55 | 0.50 | -2.40 | 1.50 | **1.39** | **0.79** |
| Religion (Muslim) | 0.97 | 0.41* | 0.43 | 0.63 | 4.73 | 2.95 | **-1.17** | **0.99** |
| Employment (farmer) | -2.30 | 0.59*** | -1.79 | 0.71* | 3.77 | 2.30 | **-2.01** | **1.18** |
| Employment (student) | 0.49 | 0.73 | 0.85 | 0.93 | 3.75 | 2.93 | **-0.34** | **1.45** |
| Place of residence (Rural) | 0.83 | 0.41* | 0.41 | 0.64 | 4.73 | 2.95 | **-0.26** | **1.11** |
| Monthly income (Medium) | 0.27 | 0.30 | -0.30 | 0.47 | 4.70 | 2.10* | **-1.80** | **0.81*** |
| Monthly income (High) | 0.63 | 0.63 | 0.55 | 0.82 | 0.18 | 2.14 | **-1.85** | **1.33** |
| Diagnostic Period length (years) | -0.02 | 0.03 | -0.01 | 0.04 | -0.10 | 0.13 | **0.05** | **0.07** |
| Drug Type (fenobarbiton) | -0.29 | 0.47 | -0.18 | 0.69 | 1.99 | 3.04 | **1.08** | **1.16** |
| Drug Amount (m/g) | -0.01 | 0.001*** | -0.003 | 0.001*** | 0.01 | 0.001* | **-0.002** | **0.004** |
| Source of drug(Health Insurance) | -3.9 | 0.61*** | -2.57 | 0.83** | 0.85 | 2.37 | **-2.38** | **1.30** |
| Source of drug(Charged) | -2.15 | 0.51*** | -2.38 | 0.81** | 1.50 | 2.62 | **-1.97** | **1.25** |
| Drug miss (Yes) | 1.68 | 0.98 | 2.86 | 1.26* | -5.50 | 4.00 | **1.28** | **1.99** |
| Family history (Yes) | 2.65 | 0.49*** | 2.25 | 0.67*** | -3.48 | 1.84 | **3.06** | **1.11**** |
| Family support (Yes) | -1.3 | 0.68 | -0.58 | 0.90 | 7.16 | 3.49* | **-2.48** | **1.48** |
| Cure (Yes) | -1.7 | 0.37*** | -1.79 | 0.53*** | 2.63 | 1.96 | **-3.39** | **0.96**** |
| Service satisfaction (Yes) | 2.9 | 0.44*** | 2.67 | 0.66*** | -10.27 | 3.43** | **4.29** | **1.20**** |
| Time (hour) | 0.44 | 0.18* | 0.37 | 0.31 | -0.33 | 0.86 | **-0.06** | **0.52** |
| Reasons of attacks (Stress) | 1.34 | 0.56* | 0.71 | 0.62 | -6.10 | 2.18** | **1.77** | **1.07** |
| Reasons for the disease (Yes) | 1.22 | 0.56* | 0.26 | 0.68 | 1.09 | 1.97 | **1-0.16** | **1.06** |
| Other disease (Yes) | 0.46 | 0.51 | 1.40 | 0.76 | -8.33 | 3.15** | **1.93** | **1.18** |
| Time (follow-up) | -0.48 | 0.03*** | -0.46 | 0.04*** | 0.93 | 0.16*** | **-0.79** | **0.10**** |
| intercept1 | 0.03 | 0.03*** | 0.06 | 0.03*** | -2.89 | 3.3*** | **1.99** | **2.3**** |
| B2 | | | 0.86 | 0.24** | | | 1.8 | **0.21**** |
| Model fit (smaller is better) | | | | | | | | |
| -2log likelihood | 448.4 | | 722.5 | | 358.3 | | **296.0** | |
| AIC | 950.9 | | 778.4 | | 824.5 | | **702.1** | |
| BIC | 1050.6 | | 881.88 | | 912.1 | | **776.5** | |

***p<0.01, **p<0.05, *p<0.10, 1 intercept is the random intercept term and  B2is the dispersion parameter

## Discussion

In many research areas, most of the time theoutcome measures are counts' with a high proportion of zeros. Therefore, for this type of data sets, ordinary least squares (OLS) regression is usually not appropriate. Recently, models such as zero-adjusting, specially zero-inflated Poisson, models have become increasingly popular for analyzing the data with count outcomes and excess zeros. Nowadays, most of the researchers use zero-adjusting Poisson models when there are extra zeros in the outcome variable (count data). This may not work for our data since we have 61 % zero values, implying that one of the listed zero-adjusting count models might be appropriate. The likelihood value statistic (−2LL) and the information criteria we used have been widely accepted for model comparison in the overall generalized linear mixed models (GLMM),including the special type of GLMM zero-adjusting models, and, for this study, zero-inflated negative binomial model was the best. Accordingly, the main important aim of this paper was modeling the number of seizure attacks (repeatedly counted) with many zeros, which is more complex than cross-sectional data due to cross-time correlations and attrition.

Age of a patient was positively and significantly associated with number of seizure attacks. This study is similar to reports from Ethiopia[32], China[33] and other countries[34,35]. In our study sex (being male) and family history of epilepsy were positively associated with the number of seizure attacks which is in line with the study[36] that reported being male is a risk factor for episode of  febrile seizures.

The present study revealed that lower socio economic status of patients was associated with the number of seizure attacks which is similar to studies conducted by different scholars  such as[37], the study suggested that low socio-economic status, indexed by low education or lack of home ownership, was a risk factor for epilepsy in adults. Other scholars also pointed out that people with epilepsy in developing regions carry a havy burden of stigma associated with poor social and economic status[38].

Moreover, in those patients having previous history of seizures and and co-infected with other disease (s), there is an increment in progression of seizures as compared to non co-infected ones.

## Conclusion and recommendation
To our knowledge, this study was the first study on statistical models for longitudinal zero-inflated count outcomes data with the application of number of seizure attacks. In this study different single level count regression models were used. Among these, the zero-inflated negative binomial regression model better fitted the data than any other models. Therefore, it is selected as the best parsimonious model to predict the number of seizure attacks. Based on our study, there is strong evidence that parameters such as age, sex, monthly income, family history of epilepsy, service satisfaction, and the presence of other diseases and time of follow-ups are considered to be the risk factors for the number of seizure attacks. As age increases, the seizure episodes also increase. As a result, those patients having seizure episodes should visit the health facilities when their age increases and those patients having seizure episodes with low income should be supported by families or the government in order to decrease the progression of the number of seizure attacks. Moreover, those patients having seizure and co-infected with other diseases should be diagnosed regularly and the underlingcauses of disease (s) should be treated in order to decrease the progression of seizure attacks.

## Strength and limitation
Longitudinally collecting data on epilepticseizure attack is really a difficult task:- hence, we were restricted us in recording the targeted sample. The small sample size is one major limitation of our study. As a result, the conclusion may not be generalized for the whole Ethiopia. Certainly, this finding would be a great help for policy-makers as baseline guide for epileptic seizure attack prevention in Ethiopia.

## Declarations
### Ethics approval and consent to participate
This study was conducted according to the principles mentioned expressed in the Declaration of Felege Hiwot Referral Hospital and Bahir Dar University, Ethiopia. It was approved by the research ethics committee at Bahir Dar University, College of Science, and all subjects who agreed to participate in this study signed a consent form.

### Consent for publication
Participants were made aware in the consent form that the results of the study would be published but that no data would be presented to allow the identification of individuals.

### Competing interests
The authors declare that they have no conflict of interests.

### Authors' Contributions
HM wrote the proposal and participated in data collection, analysis of the data and manuscript writing. DM, DT and PK approved the proposal with some revisions, participated in data collection,and analysis and manuscript writing and SPK finalized the manuscript. All authors read and approved the final manuscript.

## References
1. Carpio A, Bharucha N, Jallon P, Beghi E, Campostrini R, Zorzetto S, et al. Mortality of epilepsy in developing countries. *Epilepsia.* 2005;46(s11):28-32.doi.org/10.1111/j.1528-1167.2005.00404.x
2. Birbeck G, Chomba E, Atadzhanov M, Mbewe E, Haworth A. The social and economic impact of epilepsy in Zambia: a cross-sectional study. *The Lancet Neurology.* 2007;6(1):39-44.doi: 10.1016/S1474-4422(06)70629-9
3. Scott RA, Lhatoo SD, Sander JW. The treatment of epilepsy in developing countries: where do we go from here? *Bulletin of the World Health Organization.* 2001;79(4):344-51.
4. Organization WH. Atlas: country profiles of mental health resources 2001. 2001.

5. Tekle-Haimanot R, Forsgren L, Abebe M, Gebre-Mariam A, Heijbel J, Holmgren G, et al. Clinical and electroencephalographic characteristics of epilepsy in rural Ethiopia: a community-based study. *Epilepsy Research*. 1990;7(3):230-9.doi.org/10.1016/0920-1211(90)90020-V

6. Matuja W. Aetiological factors in Tanzanian epileptics. *East African Medical Journal*. 1989;66(5):343-8.

7. Wang P, Puterman ML. Analysis of Longitudinal Data of Epileptic Seizure Counts—A Two-State Hidden Markov Regression Approach. *Biometrical Journal*. 2001;43(8):941-62.doi.org/10.1002/1521-4036(200112)43:8<941::AID-BIMJ941>3.0.CO;2-

8. Schwebel DC, Brezausek CM. How do mothers and fathers influence pediatric injury risk in middle childhood? *Journal of Pediatric Psychology*. 2010;35(8):806-13.doi.org/10.1093/jpepsy/jsp130

9. Winckler MIB, Rotta NT. Prognostic factors for recurrence of a first seizure during childhood. *Arquivos de neuro-psiquiatria*.1997;55(4):749-56.dx.doi.org/10.1590/S0004-282X1997000500011

10. SCOTONI AE, GUERREIRO MM, ABREU HJD. First seizure: analysis of risk factors for recurrence. *Arquivos de neusro-psiquiatria*. 1999;57(2B):392-400.

11. Berg AT. Risk of recurrence after a first unprovoked seizure. *Epilepsia*. 2008;49:13-8 PubMed .doi.org/10.1111/j.1528-1167.2008.01444.x

12. Banerjee PN, Filippi D, Hauser WA. The descriptive epidemiology of epilepsy—a review. *Epilepsy Research*. 2009;85(1):31-45. PubMed doi.org/10.1016/j.eplepsyres.2009.03.003

13.Kramer MA, Cash SS. Epilepsy as a disorder of cortical network organization. *The Neuroscientist*. 2012;18(4):360-72. PubMed doi.org/10.1177/1073858411422754

14. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*. 1992;34(1):1-14. PubMed

15. Scott Long J. Regression models for categorical and limited dependent variables. *Advanced Quantitative Techniques in the Social Sciences*. 1997;7.

16. Mullahy J. Specification and testing of some modified countdatamodels.*Journalof Econometrics*.1986;33(3):341-65. doi.org/10.1016/0304-4076(86)90002-3

17.Walters GD. Using Poisson class regression to analyze count data in correctional and forensic psychology: A relatively old solution to a relatively new problem. *Criminal Justice and Behavior*. 2007;34(12):1659-74.doi.org/10.1177/0093854807307030

18.Hur K, Hedeker D, Henderson W, Khuri S, Daley J. Modeling clustered count data with excess zeros in health care outcomes research. *Health Services and Outcomes Research Methodology*. 2002;3(1):5-20.doi.org/10.1023/A:1021594923546

19. Min Y, Agresti A. Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*. 2005;5(1):1-19.doi.org/10.1191/1471082X05st084oa

20. Wolfinger R, O'connell M. Generalized linear mixed models a pseudo-likelihood approach. *Journal of statistical Computation and Simulation*. 1993;48(3-4):233-43.doi.org/10.1080/00949659308811554

21. Lawless JF. Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*. 1987;15(3):209-25. doi.org/10.2307/3314912

22. Atkins DC, Gallop RJ. Rethinking how family researchers model infrequent outcomes: a tutorial on count regression and zero-inflated models. *Journal of Family Psychology*. 2007;21(4):726.

23. Zeger SL, Liang KY. An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*.1992;11(14-15):1825-39.doi.org/10.1002/sim.4780111406

24. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*. 1993;88(421):9-25.doi.org/10.1080/01621459.1993.10594284

25. Tuerlinckx F, Rijmen F, Verbeke G, Boeck P. Statistical inference in generalized linear mixed models: A review. British Journal of Mathematical and Statistical Psychology. 2006;59(2):225-55.doi.org/10.1348/000711005X79857

26. Lin X, Breslow NE. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*. 1996;91(435):1007-16.

27.Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. Longitudinal data analysis: CRC press; 2008.

28.Wolfinger RD, Tobias RD, Sall J, editors. Mixed models: a future direction. Proceedings of the Sixteenth Annual SAS Users Group Conference, SAS Institute Inc, Cary, NC; 1991.

29.Bae S, Famoye F, Wulu J, Bartolucci AA, Singh KP. A rich family of generalized Poisson regression models with applications. *Mathematics and Computers in Simulation*. 2005;69(1):4-11.doi.org/10.1016/j.matcom.2005.02.026

30. Xie H, Tao J, McHugo GJ, Drake RE. Comparing statistical methods for analyzing skewed longitudinal count data with many zeros: An example of smoking cessation.

*Journal of Substance Abuse Treatment.* 2013;45(1):99-108.doi.org/10.1016/j.jsat.2013.01.005

31.Buu A, Li R, Tan X, Zucker RA. Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statistics in Medicine.* 2012;31(29):4074-86.doi.org/10.1002/sim.5510

32. Getachew H, Dekema N, Awol S, Abdi A, Mohammed M. Medication adherence in epilepsy and potential risk factors associated with non adherence in tertiary care teaching hospital in southwest Ethiopia. *Gaziantep Med J.* 2014;20(1):1. DOI: 10.5455/GMJ-30-45904

33. Liu J, Liu Z, Ding H, Yang X. Adherence to treatment and influencing factors in a sample of Chinese epilepsy patients. *Epileptic Disorders.* 2013;15(3):289-94.

34.Jeste SD, Patterson TL, Palmer BW, Dolder CR, Goldman S, Jeste DV. Cognitive predictors of medication adherence among middle-aged and older outpatients with schizophrenia. *Schizophrenia Research.* 2003;63(1-2):49-58.

doi.org/10.1016/S0920-9964(02)00314-6

35. Benner JS, Glynn RJ, Mogun H, Neumann PJ, Weinstein MC, Avorn J. Long-term persistence in use of statin therapy in elderly patients. *JAMA.* 2002;288(4):455-61. doi:10.1001/jama.288.4.455

36.Sharawat IK, Singh J, Dawman L, Singh A. Evaluation of risk factors associated with first episode febrile seizure. *Journal of clinical and diagnostic research: JCDR.* 2016;10(5):SC10.doi: 10.7860/JCDR/2016/18635.7853

37. Hesdorffer DC, Tian H, Anand K, Hauser WA, Ludvigsson P, Olafsson E, et al. Socioeconomic status is a risk factor for epilepsy in Icelandic adults but not in children. *Epilepsia.* 2005;46(8):1297-303.doi.org/10.1111/j.1528-1167.2005.10705.x

38. Baskind R, Birbeck GL. Epilepsy-associated stigma in sub-Saharan Africa: the social landscape of a disease. *Epilepsy & Behavior.* 2005;7(1):68-73.doi.org/10.1016/j.yebeh.2005.04.009