Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Wang et al., J Biomet Biostat 2014, 5:3
DOI: 10.472/2155-6180.1000195

Research Article Open Access

Piecewise Negative Binomial Regression in Analyzing Hypoglycemic Events with Missing Observations

Ming Wang[1]*, Junxiang Luo[2], Haoda Fu[2] and Yongming Qu[2]

[1]*Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, College of Medicine, Pennsylvania State University, Hershey, PA, USA*
[2]*Department of Biometrics, Eli Lilly and Company, Indianapolis, Indiana, USA*

Abstract

In diabetes clinical trials, hypoglycemia can be captured. Negative binomial regression is emerging as a standard method for analyzing hypoglycemic events by considering overdispersion. However, in negative binomial regression for hypoglycemic events, variability of the subjects lost to follow up due to dropout is adjusted through an offset parameter, which assumes that dropout is missing completely at random and constant hypoglycemia rate over time. This assumption is vulnerable because dropout may be due to the excessive observed hypoglycemic events and the hypoglycemic event rate may change over time. In addition, the traditional way of using negative binomial regression to analyze hypoglycemic events only compares the counts of hypoglycemic events during a specified period. However, researchers may be interested in comparing hypoglycemic event rates between treatment groups at different time periods to understand the trend over time. Fitting a negative binomial model for each time period ignoring data from other periods may decrease testing power and introduce bias if the assumption of missing completely at random does not hold. We propose piecewise negative binomial regression to incorporate multiple time periods in one model through a generalized linear mixed-effect model. Due to clinical interest, we considered multiple weighting methods to estimate the overall relative rate of hypoglycemia over multiple periods between treatments. Simulations showed that piecewise negative binomial regression performed better than the traditional negative binomial regression in preserving Type I error. As an illustration, piecewise negative binomial regression was implemented in analyzing real data from a Type 2 diabetes clinical trial.

Keywords: Hypoglycemic events; Negative binomial; Relative rate; Overdispersion; Missing at random; Counting process

Introduction

Diabetes is a chronic disease characterized by high blood glucose. Treatment for patients with Type 2 diabetes mellitus (T2DM) includes diet and exercise, oral antidiabetes agents, and injections such as insulin. The only treatment for Type 1 diabetes mellitus (T1DM) is insulin. Hypoglycemic events are common side effects of antidiabetic agents, especially insulin. It is important to develop antidiabetes agents that lead to less hypoglycemic events and better glycemic control. Therefore, it is of practical interest to use appropriate statistical methods to analyze hypoglycemic events. In clinical trials, hypoglycemic events are captured as recurrent events by patients' self-reporting. If a patient drops out of the study, hypoglycemic events and other measurements will not be recorded. As a result, missing hypoglycemia data is a common problem in diabetes clinical trials. Little and Rubin [1] defined three classes of missing data:

- Missing completely at random (MCAR): whether an observation is missing does not depend on the observed nor the unobserved values;
- Missing at random (MAR): the probability of a missing observation depends only on the observed values;
- Missing not at random (MNAR): the probability of a missing observation depends on the unobserved values.

Throughout, MAR is assumed in this research for data generation and statistical analysis of hypoglycemic events. In addition, because missing hypoglycemia is primarily due to dropout, we assume a monotone pattern of missing throughout, meaning that if a data point is missing at a specific time, the observations for this subject after that time point are also missing.

Hypoglycemic events can be treated as a count variable with the total number of events during the period of interest for each subject. Poisson and negative binomial (NB) regressions are two commonly used generalized linear models for count data [2,3]. Zero-inflated Poisson and zero-inflated NB regressions were also proposed to account for excessive zero counts [4]. Recent research demonstrated that NB regression with additional Pearson overdispersion correction and the variance-covariance of the parameters estimated through "sandwich" estimation performs the best among all the options for hypoglycemia data without missing values [5]. However, according to clinical interest, researchers may be interested in comparing hypoglycemic event rates between treatment groups at different time periods. Fitting an NB regression model at each time period separately is not optimal because the events information outside that time period is lost. Furthermore, if missing data occur under the mechanism of MAR, this method may lead to biased estimates when hypoglycemia rates are not constant over time [6]. The objective of this research is to identify a simple and effective model to analyze hypoglycemic events data in diabetes clinical trials with MAR. We propose to use piecewise negative binomial (PWNB) regression, which fits NB regression models for the count data in time intervals through a generalized linear mixed-effect model, where the time intervals are generally formed naturally based on the clinical visits or combination of multiple clinical visits. The within-subject correlation between the counts in different time

***Corresponding author:** Ming Wang, Division of Biostatistics and Bioinformatics, Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA, 17033, USA, Tel: 717-531-5745; Fax: 717-531-5779; E-mail: mwang@phs.psu.edu

Received March 31, 2014; **Accepted** May 28, 2014; **Published** May 31, 2014

Citation: Wang M, Luo J, Fu H, Qu Y (2014) Piecewise Negative Binomial Regression in Analyzing Hypoglycemic Events with Missing Observations. J Biomet Biostat 5: 195. doi:10.4172/2155-6180.1000195

intervals can be modeled through the generalized linear mixed-effect model framework. In this paper, PWNB regression will be compared with NB regression through Monte Carlo simulations and real clinical data analyses.

Methods

In this Method, we briefly describe two models: NB regression and PWNB regression. In the first model, the total count of hypoglycemia for each subject during the follow-up is summarized, and analyzed by an NB regression model. For the second model, the count of hypoglycemic events at each time interval is calculated and used for fitting longitudinal NB regression. Next, we provide the following notations:

Let N denote the total number of subjects, D_{ij} represent the j^{th} hypoglycemic event time since randomization, and C_i be the censoring time, which is the minimum of dropout time and follow-up time for the i^{th} subject. The total duration can be partitioned into m intervals by prespecified time points $0=\eta_0<\eta_1<\cdots<\eta_{m-1}<\eta_m$ where η_m is the maximum follow-up time. Note that the time intervals are determined based on the natural clinical visits and clinical interest. Assume the unit for all variables regarding time is day. Let Y_{ip} denote the number of events at time interval $[\eta_{p-1},\eta_p)$ for subject i. If $\eta_{p-1}<C_i, Y_{ip}=\sum_j I\left(\eta_{p-1}\le D_{ij}<\eta_p\right)I\left(D_{ij}<C_i\right)$, and Y_{ip} is missing if $\eta_{p-1}>C_i$, P=1,2…,m. Thus, the repeated count numbers are represented by $\boldsymbol{Y}_i^*=\left(Y_{i1},Y_{i2},\ldots,Y_{in_i^*}\right)^{\mathrm{T}}$, and the offset is $\boldsymbol{\delta}_i^*=\left(\log\frac{\eta_1-\eta_0}{30},\log\frac{\eta_2-\eta_1}{30},\ldots,\log\frac{C_i-\eta_{n_i^*-1}}{30}\right)^T$, where $n_i^*\le m$ is defined such that censor occurs at the n_i^{*th} time interval. The total number of events for the i^{th} subject is calculated by $Y_i=\sum_{p=1}^{n_i^*}Y_{ip}$ with the corresponding offset $\delta_i=\log\frac{\min\left(\eta_{in_i^*},C_i\right)}{30}$. The denominator "30" is used in the offset parameter to estimate the event rate per 30 days. If the event rate per year is to be estimated, the denominator of "365" may be used.

The dependent variable Y_i with its offset δ_i will be used for NB regression, and $\boldsymbol{Y}_i^*$ with its offset $\boldsymbol{\delta}_i^*$ will be used for PWNB regression. Let $\boldsymbol{X}_i$ denote a vector of independent variables, such as baseline covariates, treatment indicator, among others. Note that $\boldsymbol{X}_i$ also includes the period variable noted by categorical values $\left\{1,2,\ldots,n_i^*\right\}$ for PWNB regression.

Negative binomial regression

NB regression is used to model count responses, usually for overdispersed count data where the conditional variance exceeds the conditional mean. It can be considered as a generalization of Poisson regression because it has the same mean structure but an extra parameter to model the overdispersion. For instance, when subject heterogeneity in the event rate is considered, a random effect ϕ_i will be included, thus conditional on ϕ_i, $E\left(Y_i|\boldsymbol{X}_i,\ \phi_i\right)=Var\left(Y_i|\boldsymbol{X}_i,\ \phi_i\right)=\phi_i\exp\left(\beta^T\boldsymbol{X}_i\right)$, where a log link function is used to connect the mean and the linear regression of the covariate $\boldsymbol{X}_i$. If ϕ_i follows a gamma distribution with $\phi_i\sim\text{Gamma}(\kappa^{-1},\kappa)$, the marginal count observation for the i^{th} subject, $\boldsymbol{Y}_i$, follows an NB distribution defined by:

$$P(Y_i=y_i;\kappa,\mu_i)=\frac{\Gamma(y_i+\frac{1}{\kappa})}{y_i!\Gamma(\frac{1}{\kappa})}\left(\frac{\kappa\mu_i}{1+\kappa\mu_i}\right)^{y_i}\left(\frac{1}{1+\kappa\mu_i}\right)^{\frac{1}{\kappa}},\ y_i=0,1,2\ldots \quad (1)$$

Where $E\left(Y_i|\boldsymbol{X}_i\right)=\mu_i=\exp\left(\beta^T\boldsymbol{X}_i\right)$ and $Var\left(Y_i|\boldsymbol{X}_i\right)=\mu_i+k\mu_i^2$. The full-likelihood approach is used to estimate parameters. Recently, Luo and Qu [5] proposed that using "sandwich" estimation to calculate the covariance matrix of the parameter estimates together with Pearson overdispersion correction performs the most robust to model misspecification and improves the estimation efficiency by adjusting for baseline variables [6].

Piecewise negative binomial regression

PWNB regression is an extension of simple NB regression into longitudinal count data by generalized linear mixed-effect model. The normally distributed random effect is incorporated to capture the correlation of multiple counts within subjects in the analysis to improve the estimation. Following the notations in negative binomial regression, the response for the i^{th} subject is $\boldsymbol{Y}_i^*, i=1,2,\ldots,N$, and the PWNB regression model with random intercept can be written by:

$$\log\left(\mu_{ipg}\,|\gamma_i\right)=\beta_{pg}+\gamma_i \quad (2)$$

Where μ_{ipg} is the true mean event rate for subject i in treatment g (g=0 for the control group and g=1 for the treatment group) at time interval p, and β_{pg} is a scalar coefficient indicating the population-average mean event rate in treatment g at time interval p, and the random effect $\gamma_i\sim N\left(0,\sigma_\gamma^2\right)$. Of note is that we consider the simplest and commonly used form with random intercept only, but more complicated within-subject correlation can be modeled through the residuals in generalized linear model framework if necessary. Since the generalized estimating equations method may produce biased estimators under MAR assumption [7], we use pseudo-likelihood based generalized linear mixed models in estimating in the parameters in the PWNB model [8]. The "Sandwich" method is used for variance estimation. Newton-Raphson optimization technique with ridging is used to improve the likelihood of convergence [9]. Appendix A.1 and Appendix A.2 provide sample SAS codes. The relative rate of the treatment group over control group in each time interval p is:

$$\zeta_p^*=\exp\left(\beta_{p1}-\beta_{p0}\right)$$

which can be estimated directly based on the estimated parameters.

Estimate of the overall relative rate: There are three possible quantities for the overall relative rate, which will be discussed next. Given a constant relative rate over time, the unweighted overall relative rate is defined by:

$$\zeta_1=\exp\left(\bar{\beta}_{\cdot 1}-\bar{\beta}_{\cdot 0}\right) \quad (3)$$

where $\bar{\beta}_{\cdot g}=\frac{1}{m}\sum_{p=1}^{m}\beta_{pg}$. Because the time intervals may not be even, a weighted relative rate can be constructed as:

$$\zeta_2=exp\left(\sum_{p=1}^{m}w_p\beta_{p1}-\sum_{p=1}^{m}w_p\beta_{p0}\right), \quad (4)$$

with $w_p=\left(\eta_p-\eta_{p-1}\right)/\eta_m$ [10]. However, the interest of estimation may be the relative rate of the overall number of hypoglycemic events during the entire period. We can use an artificial example to illustrate the difference between the two quantities and the relative rate of the

overall events. Assume the entire period is divided into two equal intervals: in the first interval, there are 20 and 10 events for treatment groups 0 and 1, respectively; and in the second interval, there are 80 and 90 events for treatment group 0 and 1, respectively. The overall event rate ratio is 1, while the above two quantities (3) and (4) give a ratio of $\exp\left(\frac{\log 10+\log 90}{2}-\frac{\log 20+\log 80}{2}\right)=\left(\frac{10}{20}\times\frac{90}{80}\right)^{1/2}=0.75$. Therefore, we define a third relative rate with $d_p = \eta_p - \eta_{p-1}$ as:

$$\zeta_3=\frac{\sum_{p=1}^{m} d_p \mathrm{E}[\exp(\beta_{p1}+\gamma_i)]}{\sum_{p=1}^{m} d_p \mathrm{E}[\exp(\beta_{p0}+\gamma_i)]}=\frac{\sum_{p=1}^{m} d_p \exp(\beta_{p1}+\sigma_\gamma^2/2)}{\sum_{p=1}^{m} d_p \exp(\beta_{p0}+\sigma_\gamma^2/2)}=\frac{\sum_{p=1}^{m} d_p \exp(\beta_{p1})}{\sum_{p=1}^{m} d_p \exp(\beta_{p0})} \quad (5)$$

The variance for $\hat{\zeta}_3$ is estimated by the delta method:

$$\widehat{\mathrm{Var}}\left(\hat{\zeta}_3\right)=\left[\frac{\partial\hat{\zeta}_3\left(\hat{\beta}\right)}{\partial\beta}\right]'\widehat{\mathrm{Var}}\left(\hat{\beta}\right)\left[\frac{\partial\hat{\zeta}_3\left(\hat{\beta}\right)}{\partial\beta}\right] \quad (6)$$

with

$$\frac{\partial\hat{\zeta}_3\left(\hat{\beta}\right)}{\partial\beta}=\left(\frac{d_1\exp(\hat{\beta}_{11})}{\sum_{p=1}^{m} d_p\exp(\hat{\beta}_{p0})},\cdots,\frac{d_m\exp(\hat{\beta}_{m1})}{\sum_{p=1}^{m} d_p\exp(\hat{\beta}_{p0})},\frac{-d_1\exp\left(\hat{\beta}_{10}\right)\sum_{p=1}^{m} d_p\exp(\hat{\beta}_{p1})}{\left(\sum_{p=1}^{m} d_p\exp(\hat{\beta}_{p0})\right)^2},\cdots,\frac{-d_m\exp\left(\hat{\beta}_{m0}\right)\sum_{p=1}^{m} d_p\exp(\hat{\beta}_{p1})}{\left(\sum_{p=1}^{m} d_p\exp(\hat{\beta}_{p0})\right)^2}\right)'$$

The 100(1-α)% confidence interval (CI) for the $\hat{\zeta}_3$ is constructed using a log-transformation based on two reasons. First, the rate ratio estimate is a positive number; second, the log-link function is used in the PWNB model. For $\log(\hat{\zeta}_3)$, its 100(1-α)% CI can be given by:

$$\log\left(\hat{\zeta}_3\right)\pm\frac{\sqrt{\widehat{\mathrm{Var}}\left(\hat{\zeta}_3\right)}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)}{\hat{\zeta}_3} \quad (7)$$

where Φ^{-1} is the inverse function of the cumulative distribution function (CDF) of the standard normal distribution. Thus, the 100(1-α)% CI for ζ_3 is inversely calculated by an exponential transformation of the above CI in (7).

Simulation

Simulation studies are conducted to evaluate the performance of NB and PWNB models described in Methods. For each scenario, 5,000 random samples are generated and the corresponding parameters are estimated for each sample. Summary statistics based on the estimates from the 5,000 simulation samples are provided for assessing the performance of the two models.

Simulation setting

We generate recurrent hypoglycemic events through three processes: Weibull process, mixed Poisson process, and mixed inhomogeneous Poisson process [11]. Data generated from mixed Poisson process with gamma distribution will follow a negative binomial distribution, but if dropouts/missing values exist, this fact does not hold under MAR assumption. For each scenario, each treatment group includes 150 subjects and each subject will be followed up to 52 weeks.

The first generating process is Weibull, which considers the relative event rate may change over time, but assume there is no correlation among observations within a subject. The probability density function (PDF) for interevent time t_{ij} is $\lambda\gamma t_{ij}^{\gamma-1}\exp\left(-\lambda t_{ij}^{\gamma}\right)$ and CDF is $1-\exp(-\lambda t_{ij}^{\gamma})$. t_{ij} is obtained by:

$$t_{ij}=\left\{-\frac{\log u_i}{\lambda}\right\}^{1/\gamma},u_i\sim Unif\left(0,1\right) \quad (8)$$

Three possible options for the shape parameter γ include the following: $0<\gamma<1$ indicates the event rate decreases over time; $\gamma=1$ indicates the event rate stays constant over time, which leads to the Poisson process; $\gamma>1$ indicates the event rate increases over time. We set $\gamma=0.5$ for the event rate decreasing over time in an L-shape.

The second generating process is mixed Poisson, which allows for a random effect called "frailty" to be incorporated into the Poisson process to model the within-subject correlation. Conditioning on the frailty, the events are independent from each other within subject. The PDF for t_{ij} is $\lambda\phi_i\exp\left(-\lambda\phi_i t_{ij}\right)$ where $\phi_i\sim\mathrm{Gamma}(\kappa^{-1},\kappa)$ with mean 1 and variance κ. Given ϕ_i, the time t_{ij} can be simulated from independent realizations of an exponential distribution with the rate $\lambda\phi_i$:

$$t_{ij}=-\frac{\log u_i}{\lambda\phi_i},u_i\sim Unif\left(0,1\right) \quad (9)$$

For the above generating processes, we set the parameter λ to be $\lambda_C=0.4$ for the control group, and λ_T for the treatment group with the relative rate $\zeta=\frac{\lambda_T}{\lambda_C}$.

The third generating process is mixed inhomogeneous Poisson, which not only incorporates the frailty, but also allows for non-constant event rate over time. The PDF for t_{ij} is:

$$\lambda\left(t_{ij}\right)\phi_i\exp\left(-\lambda\left(t_{ij}\right)\phi_i t_{ij}\right), \quad (10)$$

where $\phi_i\sim Gamma(\kappa^{-1},\kappa)$ defined similarly in the mixed Poisson process. Here, we investigate two scenarios with piecewise event rate function. As noted in Piecewise negative binomial regression, the group indicator for the treatment group is $g=1$, otherwise, $g=0$ for the control group:

$$\lambda(t_{ij})=\begin{cases}(\alpha_0+\alpha_1 t_{ij})\exp(\beta g), & t_{ij}\le 12\\ \delta_0\exp(\beta g), & t_{ij}>12\end{cases},\ \text{where}\ \begin{cases}\alpha_0=0.15,\alpha_1=0.0208,\beta=\log(\zeta)\\ \delta_0=0.4\end{cases} \quad (11)$$

$$\lambda(t_{ij})=\begin{cases}(\alpha_0+\alpha_1 t_{ij})\exp(\beta g), & t_{ij}\le 12\\ (\delta_0+\delta_1 t_{ij})\exp(\beta g), & t_{ij}>12\end{cases},\ \text{where}\ \begin{cases}\alpha_0=0.15,\alpha_1=0.0208,\beta=\log(\zeta)\\ \delta_0=0.52,\delta_1=-0.01\end{cases} \quad (12)$$

Therefore, conditioning on ϕ_i, the event rate initially increases in the first 12 weeks, then stays stable (11) or declines (12) after 12 weeks. There are two popular methods for generating the inhomogeneous Poisson process: the inverse transform method [12] and the thinning method [13]. For easy manipulation and flexible computation, we use the latter.

To generate data for that MAR, we assume that probability of dropout at time η_p depends on the event rate prior to time η_p. The dropout probability is relative to the events that occur in the p^{th} interval:

$$\Pr\left(\text{subject } i \text{ drop out at } \eta_p\right)=f(\alpha_0+\alpha_1\left(Y_{ip}/\left(\eta_p-\eta_{p-1}\right)\right)),\ p=1,2,\ldots,m$$

Where $f(x)=1/(1+\exp(-x))$. We choose $\alpha_0=-0.6$ and $\alpha_1=5$ to generate the dropout rate ranging from 15% to 30%.

Up to this point, all simulation scenarios above assume constant relative rate over time. To investigate three estimates of overall relative rate defined in Piecewise negative binomial regression for PWNB regression, in addition to evaluating the performance under scenarios with constant relative rate over time mentioned previously, three new scenarios with non-constant relative rates over time are added to evaluate the performance of the three estimates defined in Estimate of

the overall relative rate. The new scenarios generate data from mixed inhomogeneous Poisson process (10) with κ=1, and the details of the calculations are provided in Appendix A.3:

1. "Increase+Stable":

$$\lambda(t_{ij}) = \begin{cases} \alpha_0 \exp[(\beta_1 t_{ij} + \beta_2)g], & t_{ij} \le 12 \\ \delta_0 \exp(\beta_3 g), & t_{ij} > 12 \end{cases}, \text{ where } \begin{cases} \alpha_0 = 0.8, \beta_1 = 0.1, \beta_2 = -2.5 \\ \delta_0 = 0.5, \beta_3 = -0.83 \end{cases} \quad (13)$$

The relative rate above increases over time before 12 weeks and remains constant after 12 weeks. The three quantities for the relative rate are calculated as follows: $\zeta_1 = 0.231; \zeta_2 = 0.341; \zeta_3 = 0.346.$

2. "Increase+Decrease":

$$\lambda(t_{ij}) = \begin{cases} \alpha_0 \exp[(\beta_1 t_{ij} + \beta_2)g], & t_{ij} \le 12 \\ \delta_0 \exp[(\beta_3 t_{ij} + \beta_4)g], & t_{ij} > 12 \end{cases}, \text{ where } \begin{cases} \alpha_0 = 0.8, \beta_1 = 0.1, \beta_2 = -2.5 \\ \delta_0 = 0.5, \beta_3 = -0.1, \beta_4 = 0.37 \end{cases} \quad (14)$$

The relative rate above increases over time before 12 weeks and tends to decrease after 12 weeks. The three quantities for the relative rate are calculated as follows: $\zeta_1 = 0.098; \zeta_2 = 0.073; \zeta_3 = 0.124.$

3. "Decrease+Stable":

$$\lambda(t_{ij}) = \begin{cases} \alpha_0 \exp[(\beta_1 t_{ij} + \beta_2)g], & t_{ij} \le 12 \\ \delta_0 \exp(\beta_3 g), & t_{ij} > 12 \end{cases}, \text{ where } \begin{cases} \alpha_0 = 0.8, \beta_1 = -0.1, \beta_2 = -0.1 \\ \delta_0 = 0.5, \beta_3 = -0.83 \end{cases} \quad (15)$$

The relative rate above decreases over time before 12 weeks and remains constant after 12 weeks. The three quantities for the relative rate are calculated as follows: $\zeta_1 = 0.514; \zeta_2 = 0.449; \zeta_3 = 0.466.$

For all simulations, K=5,000 Monte Carlo samples with the size of 300 (150 per group) are simulated for each scenario. The bias, Monte Carlo standard error (MCSE), mean of the estimated standard errors (SEs), mean squared error (MSE), and 95% coverage probability (CP) for log scale of the relative rate are reported and compared between NB and PWNB regressions, where the bias is calculated by $\frac{1}{K}\sum_{i=1}^{K}(\hat{\beta} - \log(\zeta))$ MCSE given by $\left(\frac{1}{K-1}\sum_{i=1}^{K}\left(\hat{\beta} - \bar{\hat{\beta}}\right)^2\right)^{1/2}$; mean of estimated SEs provided by $\frac{1}{K}\sum_{i=1}^{K} SE\left(\hat{\beta}\right)$ with $SE\left(\hat{\beta}\right)$ based on the model; MSE obtained by $\frac{1}{K}\sum_{i=1}^{K}\left(\hat{\beta} - \log(\zeta)\right)^2$. PWNB regression, the 52-week duration is divided into four intervals: 0-2, 2-12, 12-26, 26-52 week to mimic the intervals of interest for several real clinical trials. In addition, the bias, MCSE, mean of the estimated SEs, MSE, and 95% CP for the overall relative rate over all periods are compared among the three methods defined in Estimate of the overall relative rate.

Simulation results

Negative binomial vs. piecewise negative binomial regressions: Tables 1 and 2 compare the performance of the two statistical models (NB versus PWNB regression) with respect to the log scale of relative rate. For both Tables 1 and 2, the PWNB model assuming a constant relative rate over time was used in estimating the overall relative rate (i.e., the interaction between treatment and time interval was not included in the estimation model). Table 1 presents the results under the null hypothesis that there is no treatment difference. The biases from all estimates are small and comparable between the two models.

Table 1: Comparison of NB and PWNB regressions through bias, MCSE, mean of SEs, and MSE of log(risk rate) and 95% CI coverage probability (CP) based on 5,000 Monte Carlo simulations (true relative rate ζ=1.0 for treatment versus control groups).

Simulation Model	κ	Method	Bias	MCSE	Mean of SEs	MSE	CP (%)
Weibull	0	NB	0.0004	0.128	0.105	0.016	89.1
		PWNB	0.0002	0.087	0.080	0.008	94.9
Mixed Poisson	1	NBM	0.001	0.135	0.116	0.018	90.8
		PWNB	0.001	0.143	0.143	0.020	95.4
Mixed Poisson	2	NB	0.002	0.185	0.157	0.034	90.4
		PWNB	0.001	0.198	0.198	0.039	95.2
Increase+Stable	1/2	NBM	0.001	0.081	0.078	0.007	94.6
		PWNB	0.001	0.094	0.095	0.008	95.5
Increase+Stable	1	NB	-0.003	0.109	0.102	0.012	94.2
		PWNB	-0.005	0.137	0.138	0.019	95.0
Increase+Decrease	1/2	NB	0.004	0.117	0.109	0.014	93.2
		PWNB	0.003	0.107	0.108	0.011	95.1
Increase+Decrease	1	NB	-0.003	0.157	0.140	0.025	92.4
		PWNB	-0.002	0.153	0.152	0.023	94.8

Abbreviations: NB=Negative binomial; PWNB=Piecewise negative binomial; MCSE=Monte Carlo standard error; SE=Stand error; MSE=Means square error; CP=95% CI coverage probability; κ=Overdispersion parameter

Table 2: Comparison of NB and PWNB regressions through bias, MCSE, mean of SEs, and MSE of log(risk rate) and 95% CI coverage probability (CP) based on 5,000 Monte Carlo simulations (true relative rate ζ=0.6 for treatment versus control groups).

Simulation Model	κ	Method	Bias	MCSE	Mean of SEs	MSE	CP (%)
Weibull	0	NB	-0.389	0.131	0.108	0.169	88.7
		PWNB	-0.234	0.093	0.091	0.063	94.5
Mixed Poisson	1	NBM	0.002	0.134	0.116	0.018	92.0
		PWNB	0.010	0.131	0.130	0.017	94.6
Mixed Poisson	2	NB	-0.002	0.184	0.157	0.034	91.4
		PWNB	0.015	0.175	0.173	0.031	94.3
Increase+Stable	½	NBM	0.040	0.083	0.082	0.008	92.0
		PWNB	0.010	0.092	0.09	0.008	95.3
Increase+Stable	1	NB	0.064	0.113	0.107	0.017	89.3
		PWNB	0.015	0.135	0.134	0.018	94.8
Increase+Decrease	½	NB	-0.114	0.119	0.111	0.027	89.5
		PWNB	-0.059	0.111	0.111	0.016	94.4
Increase+Decrease	1	NB	-0.112	0.162	0.145	0.038	89.2
		PWNB	-0.048	0.153	0.154	0.026	93.9

Abbreviations: NB=Negative binomial; PWNB=Piecewise negative binomial; MCSE=Monte Carlo standard error; SE=Stand error; MSE=Means square error; CP=95% CI coverage probability; κ=Overdispersion parameter

PWNB regression preserves 95% CP, while NB regression inflates Type I error to a substantial degree, leading to smaller CP in all scenarios. PWNB regression has a smaller MSE for Weibull process and a larger MSE for mixed inhomogeneous Poisson process with "increase then decrease" risk rate, compared with NB regression. For NB regression, the mean of SEs is always smaller than the MCSE, which is probably why Type I error is inflated.

Table 2 shows the results for the case with true relative rate of 0.6. The 95% CP based on NB regression is lower than normal level for all scenarios, while PWNB regression preserves the appropriate coverage

Table 3: Comparison of different estimates of the overall relative rate through bias, MCSE, mean of SEs, and MSE of relative rate and 95% CI coverage probability (CP) based on 5,000 Monte Carlo simulations (true relative rate ζ=1.0).

Simulation Model	κ	Estimate	Bias	MCSE	Mean of SEs	MSE	CP (%)
Weibull	0	$\hat{\zeta}_1$	0.004	0.083	0.085	0.007	95.5
		$\hat{\zeta}_2$	0.003	0.089	0.092	0.008	95.3
		$\hat{\zeta}_3$	0.003	0.082	0.081	0.007	95.6
Mixed Poisson	1	$\hat{\zeta}_1$	0.013	0.146	0.145	0.021	95.2
		$\hat{\zeta}_2$	0.013	0.149	0.148	0.022	95.1
		$\hat{\zeta}_3$	0.013	0.149	0.148	0.022	95.0
Mixed Poisson	2	$\hat{\zeta}_1$	0.027	0.200	0.202	0.041	95.5
		$\hat{\zeta}_2$	0.030	0.202	0.204	0.042	95.6
		$\hat{\zeta}_3$	0.030	0.203	0.206	0.042	95.3
Increase+Stable	1/2	$\hat{\zeta}_1$	0.005	0.110	0.105	0.012	94.6
		$\hat{\zeta}_2$	0.002	0.095	0.096	0.009	94.7
		$\hat{\zeta}_3$	0.002	0.094	0.096	0.009	94.7
Increase+Stable	1	$\hat{\zeta}_1$	0.007	0.144	0.145	0.021	94.4
		$\hat{\zeta}_2$	0.007	0.138	0.137	0.019	94.8
		$\hat{\zeta}_3$	0.005	0.136	0.138	0.019	95.1
Increase+Decrease	1/2	$\hat{\zeta}_1$	0.010	0.118	0.117	0.014	94.2
		$\hat{\zeta}_2$	0.010	0.117	0.118	0.014	94.6
		$\hat{\zeta}_3$	0.006	0.108	0.109	0.012	95.3
Increase+Decrease	1	$\hat{\zeta}_1$	0.012	0.161	0.160	0.026	94.3
		$\hat{\zeta}_2$	0.013	0.161	0.161	0.026	94.6
		$\hat{\zeta}_3$	0.013	0.156	0.156	0.025	94.6

Abbreviations: NB=Negative binomial; PWNB=Piecewise negative binomial; MCSE=Monte Carlo standard error; SE=Stand error; MSE=Means square error; CP =95% CI coverage probability; κ=Overdispersion parameter

rate uniformly. PWNB regression has smaller or similar bias and MSE compared with NB regression for all scenarios except the mixed Poisson process. However, for the case of mixed Poisson process, the relative bias (bias divided by the MCSE) for PWNB regression is small (less than 10%). Overall, PWNB regression has better performance compared with NB regression.

Overall relative rates in piecewise negative binomial regression: Table 3 compares the three defined overall relative rates in PWNB regression when the interaction between treatment and time intervals are included in the estimation model. For all scenarios with constant relative rates over time, the results show the overall relative rates are similar between the three methods. All three methods preserve 95% CP. Table 4 shows the results for the cases with non-constant relative rates. For all three scenarios with non-constant relative rates over time, each method for the estimation of the overall relative rates preserves the 95% CP with regard to their perspective quantity in Appendix A.3.

Real Data Application

We applied the proposed methods to data from a 24-week, multicenter, open-label diabetes clinical trial for patients with T2DM who had been treated with basal insulin [14]. Three hundred seventy-four patients were randomly assigned to take either lispro mix 50/50 (LM, 50% insulin lispro protamine suspension and 50% lispro) or basal bolus therapy (BBT, glargine at bedtime plus mealtime insulin lispro). The comparisons between the two treatment groups for the whole treatment periods for nocturnal and total hypoglycemic events were reported previously [14]. In this real time application, we compared the number of hypoglycemic events between the two treatment groups for the titration period (0-12 week) and the maintenance period (12-24 week) using NB regression model and the proposed PWNB regression model. The logarithm of days in treatment divided by 30 was used as an offset parameter to estimate the hypoglycemic event rate per subject per 30 days. Table 5 shows the analysis results for the rate of hypoglycemic events for each treatment group for each treatment period, and the corresponding SE. The relative rate of LM versus BBT, the SE, 95% CI and the p-value are also reported for each treatment period. Note the weighting methods for ζ_1 and ζ_2 as described in Estimate of the overall relative rate were exactly the same for this example with equal time intervals. The estimates for the whole treatment period based on

Table 4: Comparison of different estimates of the overall relative rate through bias, MCSE, mean of SEs, and MSE of relative rate and 95% CI coverage probability (CP) based on 5,000 Monte Carlo simulations (Non-Constant Relative Rate).

Simulation Model	κ	Estimate	Bias	MCSE	Mean of SEs	MSE	CP (%)
Increase+Stable	1	$\hat{\zeta}_1$	-0.122	0.018	0.018	0.015	94.4
		$\hat{\zeta}_2$	-0.010	0.022	0.021	0.001	94.5
		$\hat{\zeta}_3$	-0.014	0.021	0.021	0.001	94.8
Increase+Decrease	1	$\hat{\zeta}_1$	-0.071	0.009	0.009	0.005	94.9
		$\hat{\zeta}_2$	-0.088	0.007	0.007	0.008	94.6
		$\hat{\zeta}_3$	-0.068	0.009	0.009	0.005	95.4
Decrease+Stable	1	$\hat{\zeta}_1$	0.049	0.079	0.079	0.009	94.8
		$\hat{\zeta}_2$	-0.026	0.064	0.063	0.005	94.4
		$\hat{\zeta}_3$	-0.009	0.065	0.064	0.004	94.5

Bias and MSE are calculated based on the overall relative rates defined by ζ_3, and 95% Coverage probabilities are calculated as the rejection rate for testing the null hypothesis with regard to their perspective definitions of overall relative rates.

Abbreviations: NB=Negative binomial; PWNB=Piecewise negative binomial; MCSE=Monte Carlo standard error; SE=Stand error; MSE=Means square error; CP =95% CI coverage probability; κ=Overdispersion parameter

Table 5: The estimated relative rate, SE for each group, and the relative rate, 95% CI and p-value for comparison of LM versus BBT for nocturnal and total hypoglycemic events.

			Rate/30 Days (Mean ± SE)		Relative Rate (LM versus BBT)		
Variable	**Model**	**Period**	**BBT[3]**	**LM[4]**	**Mean ± SE**	**95% CI**	**p-value**
Nocturnal Hypoglycemic events	PWNB	0-12 Week	0.42 ± 0.07	0.35 ± 0.04	1.19 ± 0.25	(0.78,1.80)	0.41
		12-24 Week	0.68 ± 0.08	0.51 ± 0.06	1.32 ± 0.22	(0.94,1.84)	0.10
		0-24 Week[1]	0.53 ± 0.07	0.43 ± 0.05	1.25 ± 0.21	(0.90,1.74)	0.18
		0-24 Week[2]	0.55 ± 0.07	0.43 ± 0.05	1.27 ± 0.21	(0.92,1.74)	0.15
	NB	0-12 Week	0.42 ± 0.07	0.35 ± 0.04	1.19 ± 0.25	(0.78,1.80)	0.42
		12-24 Week	0.66 ± 0.08	0.52 ± 0.06	1.27 ± 0.22	(0.90,1.78)	0.17
		0-24 Week	0.54 ± 0.07	0.43 ± 0.05	1.24 ± 0.20	(0.90, 1.72)	0.19
Total Hypoglycemic events	PWNB	0-12 Week	3.47 ± 0.29	4.09 ± 0.33	0.85 ± 0.10	(0.68,1.07)	0.16
		12-24 Week	5.24 ± 0.41	5.05 ± 0.37	1.04 ± 0.11	(0.84,1.28)	0.73
		0-24 Week[1]	4.26 ± 0.30	4.54 ± 0.31	0.94 ± 0.09	(0.77,1.14)	0.52
		0-24 Week[2]	4.35 ± 0.31	4.57 ± 0.31	0.95 ± 0.09	(0.79,1.16)	0.63
	NB	0-12 Week	3.48 ± 0.29	4.07 ± 0.33	0.85 ± 0.10	(0.68,1.07)	0.18
		12-24 Week	5.23 ± 0.41	5.19 ± 0.38	1.01 ± 0.11	(0.82,1.24)	0.95
		0-24 Week	4.33 ± 0.31	4.60 ± 0.32	0.94 ± 0.09	(0.77, 1.14)	0.53

[1]The rate and relative rate were estimated using the methods for ζ_1 and ζ_2 in the Estimate of the overall relative rate

[2]The rate and relative rate were estimated using the methods for ζ_3 as described in the Estimate of the overall relative rate

[3] N=171 and 163 for 0-12 week and 12-24 week, respectively;

[4] N=173 and 158 for 0-12 week and 12-24 week, respectively;

Abbreviation: BBT=basal bolus therapy (glargine at bedtime plus mealtime insulin lispro); LM=lispro mix 50/50 (50% insulin lispro protamine suspension and 50% lispro); LS Mean=Least square mean; SE=Standard error; NB=Negative binomial; PWNB=Piecewise negative binomial

Table 6: The estimated relative rate, SE, 95% CI and p-value for comparison of maintenance period versus titration period for nocturnal and total hypoglycemic events within each treatment group based on PWNB.

Variable	**Treatment**	**Relative Rate***			
		Mean	**SE**	**95% CI**	**p-value**
Nocturnal Hypoglycemic events	BBT	1.62	0.22	(1.25,2.10)	<0.001
	LM	1.46	0.19	(1.13,1.88)	<0.001
Total Hypoglycemic events	BBT	1.51	0.12	(1.29,1.76)	<0.001
	LM	1.23	0.08	(1.08,1.41)	0.002

*Relative rate of titration period (0-12 week) versus maintenance period (12-24 week)
Abbreviation: BBT=basal bolus therapy (glargine at bedtime plus mealtime insulin lispro); LM=lispro mix 50/50 (50% insulin lispro protamine suspension and 50% lispro); LS Mean=Least square mean; SE=Standard error; PWNB=Piecewise negative binomial

different methods of weighting were similar. For the titration period (0-12 week), the estimates provided by NB and PWNB were similar for both treatment groups. Since there were no missing values in the titration period, PWNB does not provide an advantage over NB for the titration period. For the maintenance period (12-24 week), the mean estimates of event rates were similar except for the total hypoglycemic events for LM treatment (5.05 for PWNB and 5.19 for NB, respectively). Due to the lower rate of missing values for BBT in the maintenance period (4.7%), the estimates between PWNB and NB were similar. However, as the rate of missing values was higher for LM in the maintenance period (8.7%), the difference in the estimates was larger. This indicates the potential advantage of PWNB over NB when the rate of missing values is higher and the missing is at random. When we compare the titration period with the maintenance period using PWNB, the event rates in the maintenance period were significantly higher for both nocturnal and total hypoglycemic events for both treatment groups shown in Table 6.

Summary and Discussion

NB regression is a standard method for analyzing hypoglycemic events, which are count data with overdispersion. When data is missing due to dropouts, NB regression, even with adjustment for the duration, may provide biased estimates and may inflate the Type-1 error. Simulation showed that under the mechanism of MAR, NB regression underestimates the SE and deflates 95% CP. PWNB regression, utilizing the generalized linear mixed model to incorporate the within-subject correlation, seems to provide estimators with little bias and preserves 95% CP under the assumption of missing at random. This is consistent with the finding that likelihood-based estimation can preserve Type I error under the mechanism of MAR. In addition, PWNB regression allows the estimation of the relative rate at various time periods within one model by simultaneously incorporating early dropout. We introduced three quantities of overall relative rate when PWNB regression is used. One can select the quantity of their interest based on the nature of the real data and parameter of interest. For hypoglycemic events data, we recommend the ratio of overall hypoglycemic events as the "true relative rate" because the hypoglycemic events that occurred early or late are of equal importance.

PWNB regression can be implemented through fitting a generalized linear mixed model (e.g. PROC GLIMMIX in SAS 9.2). We learned from the simulations that combining the following techniques can improve the performance of PWNB regression and make the model robust: 1) estimation based on maximization of subject-specific residual likelihood through pseudo-likelihood technique with Taylor linearization; 2) Newton-Raphson ridge optimization; and 3) covariance structure of estimates calculated by "Sandwich" estimation.

In simulation, we used conditional PWNB regression with a random effect as (2) in the simulation. Another way to model the within-subject correlation for the same subject is to fit a marginal model, i.e., the correlation is modeled through the residuals. The marginal model is generally hard to converge, especially for large number of periods.

For the simulation scenarios in simulation, because there are four time intervals, the convergence could not be achieved for some samples for the marginal models. Therefore, the conditional model was used in the simulation. We tested the marginal model for a simple scenario with two time intervals and found it can preserve the 95% CP well. In the real data application in real data application, the marginal model was used.

There are several limitations for this research. First, we assume the dropout occurred exactly at the end of each time period. In reality, dropout may occur at any time. More complex dropout scenarios may be explored in future research. Second, we assumed MAR in the simulation. For MNAR, other existing techniques such as pattern mixture models can be combined with PWNB regression to construct more reliable estimators. Third, we select the clinical visits as the cutoff points to divide the follow-up into intervals based on clinical interest. How to choose the appropriate thresholds and optimal number of intervals for more accurate inference from PWNB regression is still an open question. Fourth, the PWNB regression is an approximation to the true event rate, which is generally believed to be a smooth function of time although simulation shows such an approximation provides excellent results.

Acknowledgement

Part of this work was done when Dr.Ming Wang was a summer intern at Eli Lilly and Company. The authors would like to thank Dr.Qianyi Zhang for her careful scientific review. We would also like to thank the two anonymous referees for their useful comments.

References

1. Little RJ, Rubin DB (2002) Statistical Analysis with Missing Data. Wiley: New York.
2. Agresti A (2002) Categorical Data Analysis. Wiley: New York.
3. Li CS (2010) Semiparametric negative binomial regression models. Communications in Statistics-Simulation and Computation 39: 475-486.
4. Greene WH (1994) Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. NYU Working Paper No. EC-94-10.
5. Luo J, Qu Y (2013) Analysis of hypoglycemic events using negative binomial models. Pharm Stat 12: 233-242.
6. Bond SJ, Farewell VT (2009) Likelihood estimation for a longitudinal negative binomial regression model with missing outcomes. J R Stat Soc Ser C Appl Stat 58: 369-382.
7. Troxel AB, Lipsitz SR, Brennan TA (1997) Weighted estimating equations with nonignorably missing response data. Biometrics 53: 857-869.
8. Frank Liu G, Zhan X (2011) Comparisons of methods for analysis of repeated binary responses with missing data. J Biopharm Stat 21: 371-392.
9. Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53: 983-997.
10. Lambert P (1996) Modeling of repeated series of count data measured at unequally spaced times. Applied Statistics 45: 31-38.

Citation: Wang M, Luo J, Fu H, Qu Y (2014) Piecewise Negative Binomial Regression in Analyzing Hypoglycemic Events with Missing Observations. J Biomet Biostat 5: 195. doi:10.4172/2155-6180.1000195