Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2024, 2024:128

https://doi.org/10.28919/cmbn/8887

ISSN: 2052-2541

HURDLE REGRESSION MODELING: PARKINSON'S DISEASE DATA, SPECIFICALLY CLINICAL PARTICIPANTS OF MOVEMENT DISORDERS SOCIETY UNIFIED PARKINSON'S DISEASE RATING SCALE

ANISA HAURA SALSA FATIH YUSUF, NURTITI SUNUSI*, ERNA TRI HERDIANI

Department of Statistics, Hasanuddin University, Makassar 90245, Indonesia

Abstract: Parkinson's disease is a disease that attacks the motor parts of the body so that it can reduce the quality of life of sufferers. Treatment is limited to dopaminergic and physical therapy. This study aims to determine the factors suspected of influencing motor complications, especially in clinical participants of the Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) who are diagnosed with Parkinson's disease. MDS-UPDRS is an assessment instrument used to measure the severity of motor complication symptoms. Motor complication data obtained from MDS-UPDRS shows excessive distribution characteristics at excess zero values (zero inflation), which indicates overdispersion due to excess zero. Therefore, Hurdle regression is needed to overcome this problem. The predictor variables used are part of the measurement of MDS-UPDRS, namely motor aspects, non-motor aspects in daily life, and motor examinations by medical professionals. The results shows that Hurdle negative binomial regression model was better than the Hurdle Poisson and Hurdle Conway Maxwell Poisson when applied to motor complication data. Based on the Hurdle negative binomial regression model, it is known that clinical participants who do not experience motor complications are significantly influenced by non-motor aspect variables in daily life. In addition, in the count model, each increase in the score on the non-motor aspect variable tends to increase the average motor complication score by 1.0217.

Keywords: Parkinson’s disease; overdispersion; Hurdle; Hurdle negative binomial; Hurdle Conway Maxwell Poisson.

2020 AMS Subject Classification: 62P10, 62J20.

*Corresponding author

E-mail address: nurtitisunusi@unhas.ac.id

Received September 04, 2024

1. INTRODUCTION

Parkinson's disease is a neurological disorder in the brain that causes the loss of dopamine neurons in the midbrain, namely the substantia nigra, which regulates balance and motor skills. This disease usually attacks the elderly, but it is possible for it to attack young people. This can happen because of genetic and environmental factors that increase the risk of developing Parkinson's disease. The motor symptoms that usually occur are bradykinesia, uncontrollable tremor at rest, muscle stiffness, and postural impairment [1]. The treatment that can be given is by providing dopaminergic and physical therapy. However, sometimes in the treatment process, motor complications can occur. To evaluate the severity of symptoms and monitoring possible motor complications, the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) is used. This tool helps medical professionals evaluate the patient's overall condition, including motor and non-motor symptoms. The characteristics of individuals at high risk for motor complications can be evaluated using three parts of the MDS-UPDRS measurement, including non-motor aspects in daily life, motor aspects in daily life, and motor examinations carried out directly by medical professionals [2].

Therefore, a method is needed that can be used to identify factors suspected of influencing the occurrence of motor complications. The relationship between the number of motor complications and factors suspected of influencing them can be determined using the Poisson regression model. The Poisson regression model is used to model data whose response variable is count data. The assumptions are equidispersion, the average value and variance of the response variable must have the same value. However, it often happens in data when the average value and variance have different values. When the average value is smaller than the variance value, the data is overdispersion and vice versa when the average value is greater than the variance value, the data is underdispersion [3]. One possible cause of this is the presence of excess zero [4]. To overcome this, a method is needed that can overcome equidispersion violations due to excess zero, namely by using Hurdle regression. Hurdle regression consists of two model parts, namely: the logit model (logistic regression) and the count model (truncated regression) [5]. In this study, several count models will be used in Hurdle regression, namely Poisson, negative binomial, and Conway Maxwell Poisson which will be applied to the data on the number of motor complications of MDS-UPDRS clinical participants.

2. MATERIAL AND METHODS

2.1.Dispersion

Dispersion testing is done with a deviance test to check for possible equidispersion violations in

data. If the calculated deviance value divided by the degrees of freedom is greater than one, it indicates that the data is overdispersion. Conversely, if the result of dividing the deviance value by the degrees of freedom is less than one, it indicates that the data is underdispersion. The deviance test is calculated using the following equation [6]:

$$deviance = 2\sum_{i=1}^{n}\left[y_i \log\frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i)\right] \quad (1)$$

df : degrees of freedom, $df = n - (p + 1)$

y_i : value of the i[th] response variable with $i = 1,2, \dots, n$

$\hat{y}_i$: Poisson regression estimator, $\hat{y}_i = e^{x_i \beta_{Poisson}}$

2.2.Poisson Regression

If y_i is a response variable that has a Poisson distribution for each observation i, which describes the number of events with n observations, then the probability density function for y_i:

$$f(y_i; \mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \quad (2)$$

with $y = 0,1,2, \dots, n$. The natural parameters of the equation are $\ln \mu$ with the log natural canonical link function, $\eta_i = \ln \mu_i$. If the relationship between μ_i and the linear predictor η_i is expressed by $\ln(\mu_i) = \eta_i$ then the Poisson regression model is obtained as follows [6]:

$$\mu_i = \exp\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right) \quad (3)$$

2.3.Negative Binomial Distribution

The negative binomial distribution is an effective approach used to handle overdispersion in count data [7]. The negative binomial distribution has a dispersion parameter that controls the variability of the data. The probability density function is as follows [5]:

$$f(y; \mu, \phi) = \frac{\Gamma(y + \phi)}{\Gamma(y + 1)\Gamma(\phi)}\left(\frac{\mu}{\mu + \phi}\right)^{y}\left(\frac{\phi}{\mu + \phi}\right)^{\phi} \quad (4)$$

with y is the count data value, $\mu = E(y)$, and ϕ^{-1} is the dispersion parameter.

2.4.Conway Maxwell Poisson Distribution

The Conway Maxwell Poisson (CMP) distribution is a development of the Poisson distribution, which was first introduced by Conway and Maxwell. According to Sellers et al. [8] the CMP

distribution can handle data that experiences overdispersion or underdispersion. The CMP distribution controls the level of data dispersion with parameter v, namely equidispersion $(v = 1)$, overdispersion $(0 \leq v < 1)$, and underdispersion $(v > 1)$. This distribution also includes several general distributions as special cases: the Poisson distribution $(v = 1)$, the geometric distribution $(v = 0$ dan $\mu < 1)$, and the Bernoulli distribution $(v \to \infty)$. The following is the form of the CMP distribution probability density function:

$$f(y; \mu, v) = \frac{\mu^y}{(y!)^v} \frac{1}{Z(\mu; v)}, \ \mu > 0; v \geq 0 \tag{5}$$

with $Z(\mu, v) = \sum_{k=0}^{\infty} \frac{\mu^k}{(k!)^v}$ is the normalization constant, $\mu = E(y)$, and v is the dispersion parameter [9].

2.5.Hurdle Regression

One of the methods to overcome overdispersion is Hurdle regression. Overdispersion occurs when the data variance is greater than the average value. One of the causes of overdispersion is excessive zero values in the response variable (excess zero). Hurdle regression divides the model into two parts, namely: a model for binary data with zero or positive values, which is estimated using logistic regression, and a model for data with positive values, estimated with a truncated model. The Hurdle regression probability distribution is as follows [10]:

$$f(y_i) = \begin{cases} f_1(0), & y_i = 0 \\ \dfrac{1 - f_1(0)}{1 - f_2(0)} f_2(y_i), & y_i > 0 \end{cases} \tag{6}$$

$f_1(0)$ is the probability when $y_i = 0$

$f_2(y_i)$ is the probability density function of y_i when $y_i > 0$

2.6.Hurdle Poisson Regression

If $y_i = 1,2, \ldots, n$ is the n^{th} response variable from the i^{th} observation measurement, then the probability density function of the Hurdle Poisson (Hpois) regression formed based one Equation (2) and (6) is shown in Equation (7):

$$P(Y_i = y_i) = \begin{cases} (1 - \pi_i), & y_i = 0 \\ (\pi_i) \dfrac{e^{-\mu_i} \mu_i^{y_i}}{y_i! \, (1 - e^{-\mu_i})}, & y_i > 0 \end{cases} \tag{7}$$

The probability of a value other than zero occurring is defined as π_i, with $0 < \pi_i < 1$. μ_i is the mean of a Poisson distribution with $\mu_i > 0$. It is known that π_i dan μ_i depend on vector of

predictor variables, which can be defined as follows:

$$\pi_i = \frac{\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)}{1+\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)} \qquad \mu_i = \exp\left(\sum_{j=1}^{p} x_{ij}u_j\right)$$

with $0 < \pi_i < 1,\ i = 1,2,\dots,n$, and $j = 1,2,\dots,p$.

2.7.Hurdle Negative Binomial Regression

Hurdle Negative Binomial (HNB) Regression is a Hurdle regression whose second part of the model uses a negative binomial distribution. This regression is effective for data that shows overdispersion due to excessive zeros in the response variable [11]. If $y_i = 1,2,\dots,n$ is the n^{th} response variable from the i^{th} observation measurement, then the probability density function of the HNB regression formed based on Equation (4) and (6) is shown in Equation (8) [12]:

$$P(Y_i = y_i) = \begin{cases} (1-\pi_i), & y_i = 0 \\ (\pi_i)\dfrac{\Gamma(y_i+\phi)}{\Gamma(y_i+1)\Gamma(\phi)}\left(\dfrac{\mu_i}{\phi+\mu_i}\right)^{y_i}\dfrac{(1+\phi^{-1}\mu_i)^{-\phi}}{1-(1+\phi^{-1}\mu_i)^{-\phi}}, & y_i > 0 \end{cases} \tag{8}$$

The probability of a value other than zero occurring is defined as π_i, with $0 < \pi_i < 1$. μ_i is the mean of a negative binomial distribution with $\mu_i > 0$ and ϕ^{-1} is the dispersion parameter with $\phi > 0$. It is known that π_i dan μ_i depend on vector of predictor variables, which can be defined as follows:

$$\pi_i = \frac{\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)}{1+\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)} \qquad \mu_i = \exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)$$

with $0 < \pi_i < 1,\ i = 1,2,\dots,n$, and $j = 1,2,\dots,p$.

2.8.Hurdle Conway Maxwell Poisson Regression

Hurdle Conway Maxwell Poisson (HCMP) regression uses Hurdle regression with a second model based on the CMP distribution. This regression uses the v parameter to adjust the dispersion in the data so that it is flexible in dealing with overdispersion or underdispersion in the data. If $y_i = 1,2,\dots,n$ is the n^{th} response variable from the i^{th} observation measurement, then the probability density function of the HCMP regression formed based on Equation (5) and (6) is shown in Equation (9) [13]:

$$P(Y_i = y_i) = \begin{cases} (1-\pi_i), & y_i = 0 \\ (\pi_i)\dfrac{1}{Z(\mu_i;v)-1}\dfrac{\mu_i^{y_i}}{(y_i!)^v}, & y_i > 0 \end{cases} \tag{9}$$

Sellers [13] states that the function $Z(\mu_i; v)$ can be approximated by

$$Z(\mu_i, v) \approx \frac{\exp\left(v\mu_i^{\frac{1}{v}}\right)}{\mu_i^{\frac{v-1}{2v}}(2\pi^*)^{\frac{v-1}{2}}\sqrt{v}} \tag{10}$$

It is known that π_i dan μ_i depend on vector of predictor variables, which can be defined as follows:

$$\pi_i = \frac{\exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)}{1 + \exp\left(\sum_{j=1}^{p} z_{ij}\delta_j\right)} \qquad \mu_i = \exp\left(\sum_{j=1}^{p} x_{ij}\gamma_j\right)$$

with $i = 1,2, \dots, n$, and $j = 1,2, \dots, p$.

2.9.Algorithm of Broyden-Fletcher-Goldfarb-Shanno

Algorithm of *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) is a development of the Newton-Raphson method. The main difference lies in the calculation of the hessian matrix. In Newton-Raphson, the Hessian matrix is calculated directly from the second derivative of the log-likelihood function. Meanwhile, in BFGS, the Hessian matrix is calculated iteratively using information obtained from each iteration step. Parameter estimates are obtained with the following equation [14]:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \lambda_k \boldsymbol{H}_k^{-1}\boldsymbol{g}_k \tag{11}$$

with

$\lambda_k = \min \ln L(\boldsymbol{\theta}_k - \lambda \boldsymbol{H}_k^{-1}\boldsymbol{g}_k)$

$\boldsymbol{g}$: the first derivative of log-likelihood function$\left(\nabla f(\boldsymbol{\theta}_k)\right)$.

$\boldsymbol{H}$: the Hessian matrix

$$\boldsymbol{H}_{k+1} = \left(\boldsymbol{I} - \frac{\boldsymbol{s}_k \boldsymbol{y}_k^T}{\boldsymbol{y}_k^T \boldsymbol{s}_k}\right)\boldsymbol{H}_k\left(\boldsymbol{I} - \frac{\boldsymbol{y}_k \boldsymbol{s}_k^T}{\boldsymbol{y}_k^T \boldsymbol{s}_k}\right) + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^T}{\boldsymbol{y}_k^T \boldsymbol{s}_k} \tag{12}$$

$\boldsymbol{s}_k = \boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k$

$\boldsymbol{y}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$

$\boldsymbol{H}_0 =$ the identity matrix $n \times n$

The iteration process will continue until the parameter estimates converge, namely when $\|\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k\| \leq \varepsilon$, where ε is a very small number close to zero.

2.10. Akaike Information Criteration

One of the important tools for selecting the best model is to use the Akaike Information Criteria (AIC). AIC assesses how well the model represents the data by considering the number of parameters and the degree of fit to the data. Models with lower AIC values are considered better. The AIC difference between two models indicates how much one model is better than the other. The AIC value can be defined as:

$$AIC = -2\log L(\hat{\theta}) + 2p \tag{13}$$

where $L(\hat{\theta})$ is the likelihood value of the model given the data and p is the number of parameters in the model [15].

3. MAIN RESULTS

3.1 Descriptive Data

The data used are secondary data obtained from the Parkinson's Progression Markers Initiative (PPMI) website, updated version of April 17, 2024, accessed on April 18, 2024 [16]. The number of observations consist of 362 clinical participants. The secondary data in question is MDS-UPDRS clinical data, namely the number of non-motor aspect scores in daily life, the number of motor aspect scores in daily life, the number of motor examination scores carried out by medical professionals, and the number of motor complication scores. The variables used a Likert scale with a score range of 0-4. A higher score indicates a more severe symptom impact. An overview of MDS-UPDRS clinical participants by age and gender is shown in Table 1.

Table 1. Frequency by Age and Gender

Age	Gender				Total	
	Male		Female			
	Frequency	%	Frekuensi	Frequency	%	%
Age < 50th	11	3.04	4	1.10	15	4.14
50th ≤ Age ≤ 59th	43	11.88	34	9.39	77	21.27
60th ≤ Age ≤ 69th	77	21.27	49	13.54	126	34.81
Age ≥ 70th	98	27.07	46	12.71	144	39.78
Total	229	63.26	133	36.74	362	100.00

Table 1 shows that in the MDS-UPDRS clinical participants, the most based on gender category were male with a percentage of 63.26% while female were 36.74%. Then based on age category, the highest frequency in MDS-UPDRS clinical participants was in the age group $60\text{th} \leq \text{Age} \leq 69\text{th}$ and $\text{Age} \geq 70\text{th}$. This shows that the highest frequency occurs in male with the age group $\geq 70\text{th}$.

Furthermore, in this study, an overdispersion test was conducted with a deviance test on the response variable, namely the number of motor complications. Based on the deviance test, a dispersion value of 3.733 was obtained which was greater than one. This indicates that the data on the number of motor complications indicates overdispersion. In addition, based on Figure 1, it can be seen that most of the data is concentrated on a score of zero. This can indicate that a high level of motor complications is relatively rare. The graph also shows an excess zero in the variable number of motor complication scores, where the frequency of zero values is much greater than other values. Therefore, the analysis can be continued by applying the Hurdle regression model. Model analysis is carried out with the help of the BFGS algorithm.

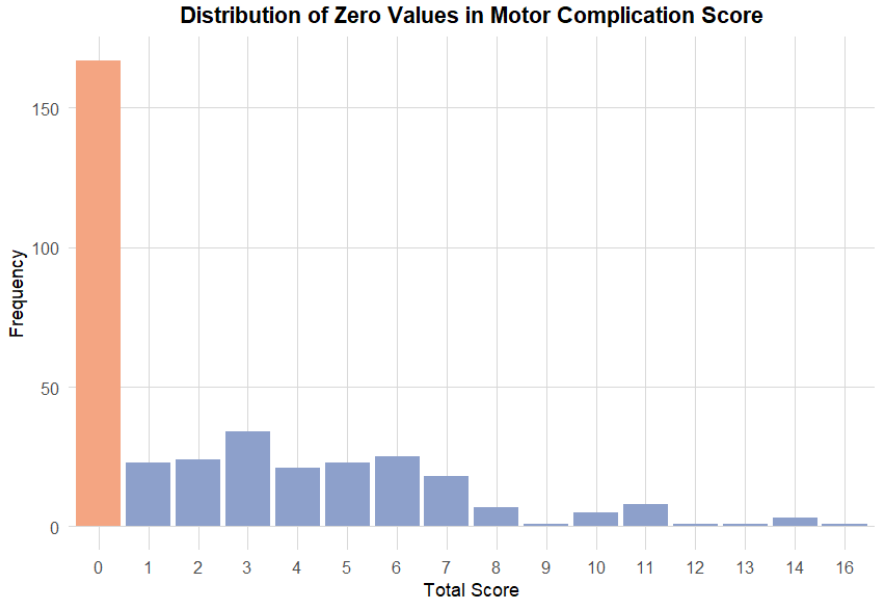

Figure 1. Distribution of Zero Values in Motor Complication Score

3.2 Modeling Hurdle Regression

Data modeling of the number of motor complications in MDS-UPDRS clinical participants was performed using the HPois, HNB, and HCMP regression models with the BFGS algorithm approach. The estimation results of each parameter are shown in Table 2:

Table 2. Hurdle Regression Parameter Estimation

Models		Parameter	Estimate	*Std.error*	*Wald test* (W)
Hurdle Poisson	Logit	δ_0	-0.5950	0.2910	-
		$\boldsymbol{\delta_1}$*	**0.0811**	**0.0245**	**10.9470**
		δ_2	0.0252	0.0236	1.1462
		δ_3	-0.0097	0.0093	1.0929
	Poisson Count	u_0	1.1846	0.0861	-
		$\boldsymbol{u_1}$*	**0.0215**	**0.0066**	**10.5875**
		$\boldsymbol{u_2}$*	**0.0138**	**0.0069**	**3.9644**
		u_3	-0.0018	0.0025	0.4963
Hurdle Negative Binomial	Logit	δ_0	-0.5967	0.2911	-
		$\boldsymbol{\delta_1}$*	**0.0812**	**0.0245**	**10.9654**
		δ_2	0.0252	0.0236	1.1434
		δ_3	-0.0096	0.0093	1.0831
	Negative Binomial Count	β_0	1.1545	0.1204	-
		$\boldsymbol{\beta_1}$*	**0.0215**	**0.0096**	**5.0477**
		β_2	0.0142	0.0100	2.0454
		β_3	-0.0019	0.0035	0.3064
		ϕ	1.6383	0.2687	37.1677
Hurdle Conway Maxwell Poisson	Logit	δ_0	-0.5968	0.2911	-
		$\boldsymbol{\delta_1}$*	**0.0812**	**0.0245**	**10.9660**
		δ_2	0.0252	0.0236	1.1434
		δ_3	-0.0096	0.0093	1.0829
	Conway Maxwell Poisson Count	γ_0	0.7180	0.1192	-
		$\boldsymbol{\gamma_1}$*	**0.0140**	**0.0055**	**6.4941**
		γ_2	0.0089	0.0056	2.5248
		γ_3	-0.0012	0.0020	0.3366
		$\exp(v)$	0.6367	0.1122	16.1972

*) indicates that variables have significant affect

Each Hurdle model consists of two parts, namely the logit model and the count model. The logit model for each HPois, HNB, and HCMP regression has almost the same parameter estimation results. In addition, in the logit model based on the Wald test value presented in Table 2, a significant variable is obtained, namely the variable number of non-motor aspect scores in daily life (X_1) with a Wald test value of $W > \chi^2_{(0.05,1)} = 3.841$. This shows that the chances of clinical

participants not experiencing motor complications are significantly influenced by the variable number of non-motor aspect scores in daily life.

As for the count model for each regression of HPois, HNB, and HCMP, the results are quite different. The Poisson count model has two variables that significantly affect the variable number of motor complication scores, including the variable number of non-motor aspects (X_1) and motor in daily life (X_2) with a Wald test value of $W > \chi^2_{(0.05,1)} = 3.841$. This means that for every one unit increase in non-motor scores in daily life (X_1), the average number of motor complication scores will increase by $\exp(0.0215) = 1.0217$ assuming other variables are constant, and for every one unit increase in motor scores in daily life (X_2), the average number of motor complication scores will increase by $\exp(0.0138) = 1.0139$ assuming other variables are constant.

The negative binomial count and Conway Maxwell Poisson count models have one significant variable respectively that affects the variable of the number of motor complication scores, namely the number of non-motor aspect scores in daily life (X_1) with Wald test value of $W > \chi^2_{(0.05,1)} = 3.841$. This means that in the negative binomial count model for every one unit increase in non-motor score in daily life, the average number of motor complication scores will increase by $\exp(0.0215) = 1.0217$ assuming other variables are constant, while in the Conway Maxwell Poisson count model for every one unit increase in non-motor score in daily life, the average number of motor complication scores will increase by $\exp(0.0140) = 1.0141$ assuming other variables are constant.

3.3 Model Comparison

The selection of the best model to apply the regression model to MDS-UPDRS clinical data will be tested by selecting the minimum Akaike Information Criterion (AIC) value. The respective AIC values for the HPois, HNB, and HCMP regression models are shown in Table 3.

Table 3. AIC Value of Hurdle Regression Model

Regression Model	AIC Value
HPois	1434.676
HNB	1386.599
HCMP	1409.921

Table 3 shows that the HNB regression model has a smaller AIC value than the HPois and HCMP

regression models. This indicates that the negative binomial Hurdle regression model is better to be applied to MDS-UPDRS clinical data that is overdispersion with excess zero.

In summary, the analysis of the logit and count components within the Hurdle Poisson (HPois), Hurdle Negative Binomial (HNB), and Hurdle Conway Maxwell Poisson (HCMP) regression models indicates that the HNB model is the most appropriate for modeling MDS-UPDRS clinical data characterized by overdispersion and an excess of zero counts. The logit models across all three regressions exhibit consistent results, identifying non-motor aspect scores in daily life (X_1) as a significant predictor of the probability that clinical participants will not experience motor complications. In contrast, the count models demonstrate variability, with the HNB model outperforming others based on the Akaike Information Criterion (AIC). This outcome corroborates prior research by Sellers [13], who also found that the negative binomial model was superior to the Conway-Maxwell Poisson model for handling data with overdispersion due to excess zeros, thereby reinforcing the suitability of the HNB model in this analytical context.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] J.C. Bridi, F. Hirth, Mechanisms of α-Synuclein induced synaptopathy in Parkinson's disease, Front. Neurosci. 12 (2018), 80. https://doi.org/10.3389/fnins.2018.00080.

[2] C.G. Goetz, B.C. Tilley, S.R. Shaftman, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results, Move. Disord. 23 (2008), 2129–2170. https://doi.org/10.1002/mds.22340.

[3] P. Schober, T.R. Vetter, Count data in medical research: Poisson regression and negative binomial regression, Anesth. Analg. 132 (2021), 1378–1379. https://doi.org/10.1213/ane.0000000000005398.

[4] D.B. Bekalo, D.T. Kebede, Zero-inflated models for count data: An application to number of antenatal care service visits, Ann. Data. Sci. 8 (2021), 683–708. https://doi.org/10.1007/s40745-021-00328-x.

[5] N. Bhaktha, Properties of hurdle negative binomial models for zero-inflated and overdispersed count data, Thesis, The Ohio State University, 2018.

[6] A. Agresti, Categorical data analysis, Wiley, 2002. https://doi.org/10.1002/0471249688.

[7] M.R. Abonazel, S.M. El-sayed, O.M. Saber, Performance of robust count regression estimators in the case of overdispersion, zero inflated, and outliers: simulation study and application to German health data, Commun.

Math. Biol. Neurosci. 2021 (2021), 55. https://doi.org/10.28919/cmbn/5658.

[8] K.F. Sellers, S. Borle, G. Shmueli, The COM-Poisson model for count data: a survey of methods and applications, Appl. Stoch. Models Bus. Ind. 28 (2012), 104-116. https://doi.org/10.1002/asmb.918.

[9] K.F. Sellers, B. Premeaux, Conway-Maxwell-Poisson regression models for dispersed count data, WIREs Comput. Stat. 13 (2020), e1533. https://doi.org/10.1002/wics.1533.

[10] J. Mullahy, Specification and testing of some modified count data models, J. Econometrics 33 (1986), 341–365. https://doi.org/10.1016/0304-4076(86)90002-3.

[11] C.D. Desjardins, Evaluating the performance of two competing models of school suspension under simulation - The zero-inflated negative binomial and the negative binomial hurdle, Thesis, The University of Minnesota, 2013.

[12] S.E. Saffari, R. Adnan, W. Greene, Hurdle negative binomial regression model with right censored count data, SORT: Stat. Oper. Res. Trans. 36 (2012), 181–194.

[13] K.F. Sellers, The Conway–Maxwell–Poisson distribution, Cambridge University Press, 2023.

[14] N. Ullah, J. Sabi’ u, A. Shah, A derivative-free scaling memoryless Broyden-Fletcher-Goldfarb-Shanno method for solving a system of monotone nonlinear equations, Numer. Linear Algebra Appl. 28 (2021), e2374. https://doi.org/10.1002/nla.2374.

[15] P.C. Emiliano, M.J.F. Vivanco, F.S. de Menezes, Information criteria: How do they behave in different models?, Comput. Stat. Data Anal. 69 (2014), 141–153. https://doi.org/10.1016/j.csda.2013.07.032.

[16] Parkinson’s Progression Markers Initiative, Study data. https://ida.loni.usc.edu/pages/access/studyData.jsp?project=PPMI, Accessed: Apr. 18, 2024.