Modeling Number of Tetanus Neonatorum Cases Using Generalized Poisson Regression and Zero Inflated Generalized Poisson

Selvia Nurul Qomari [1*], Bambang Widjanarko Otok[2], Hari Basuki N[1], Arief Wibowo[1]

[1] Department of Biostatistics and Demography, Faculty of Public Health, Airlanga University, Surabaya, (Indonesia)
[2] Laboratory of Environmental and Health Statistic, Department of Statistic, Institut Teknologi Sepuluh Nopember (ITS), Surabaya (Indonesia)

ABSTRACT

Tetanus neonatorum is one of the main causes of death in newborns. The infant mortality rate due to tetanus neonatorum in 2015 in Indonesia is 50.9%. East Java in 2015, there were 13 babies who died from tetanus neonatorum or about 59.9% of the total 22 cases of tetanus neonatorum that occurred. Data on the number of tetanus neonatorum in East Java is the data count and has a value of zero excess. Therefore, the purpose of this study is to examine the factors that can lead to neonatorum tetanus cases using Generalized Poisson regression and Generalized Zero-inflated Poisson, so it is expected to reduce the infant mortality rate (IMR) due to a bacterial infection. The results showed that ZIGP regression modeling gave the best results based on Akaike Information Criterion (AIC). Regression model for count model, the number of tetanus neonatorum cases is influenced by percentage of pregnant woman K4, percentage of immunization status of TT2 + pregnant women, percentage of shaman's birth, while in zero-inflated model only influenced by percentage of immunization status of TT2 + pregnant women. The decline in the average number of cases of neonatal tetanus in a region affected by the percentage of pregnant women visit K4 and the percentage of TT2 + status of pregnant women increased, while the percentage of deliveries shaman is high will increase the average number of cases of neonatorum tetanus.

KEYWORDS: Tetanus neonatorum, Data Count, GP, ZIGP, AIC

INTRODUCTION

Tetanus Neonatorum typically develops within the first or second week of infant life and is often referred to as seventh or eighth day illness [1], and can bring deaths in 70-90% of cases. Modern medical treatments, which are rare in the third world where the disease is very prevalent, rarely reduce mortality to less than 50% [2].

WHO estimates that in 2008, 59,000 newborns died from TN, a decline of 92% from the situation in the late 1980s. In 2008 there were 46 countries that still have not eliminated Tetanus Maternal and Neonatal (TMN), one of them is Indonesia. Prior to the introduction of tetanus neonatorum (TN) elimination efforts, Indonesia is one of the countries with the highest rates in Asia. Community-based surveys for TN deaths were conducted in the early 1980s in Jakarta and rural areas in Bali, Java, Kalimantan, West Nusa Tenggara, East Nusa Tenggara, Sumatra, and Sulawesi revealed mortality rates ranging from 6-23 TN deaths per 1000 live births [3].

Research on risk factors associated with tetanus neonatorum has been done [4]. The results obtained from the research are information on several risk factors that are significantly related to tetanus neonatorum incidence which include sterility factor of umbilical cord handling, umbilical cord handling factor, TT immunization status factor, birth attendant factor, and pregnancy check factor. In 2015, there were 22 cases of tetanus neonatorum in East Java with the most cases found in Bangkalan which accounted for 8 cases. However, there are no TN cases in the regency / city, in other words the TN case in the district / city is worth 0 (zero), i.e. 27 out of 38 districts / cities in East Java or about 71% [5]. The proportion of 71% value of zero indicates that in the case of TN in East Java in 2015 had a zero-inflation and tends to over dispersion where the mean value is 0.5789 while the larger variance namely 2,142. This leads to the assumption over dispersion condition Poisson regression can't be met [6]. Over dispersion also will have consequences on the value of the standard error estimator for a smaller (underestimate) which in turn will lead to an error in the inference for the parameters [7].

Some studies related to count data modeling, [8] analyzed the cases of domestic violence, that [9][10] the data modeling over dispersion Poisson suffered because of the value zero with Zero-inflated Poisson (ZIP) is not accurate enough, while [11] using ZINB not assessed in accordance with the existing data distribution. [12] which model the factors affecting maternal mortality in East Java by using ZIP Regression, that the modeling data that contains lots of zeros by using ZIP provides better results when compared to using linear and poisson regression a criteria AIC [13].

The above description shows that the number of tetanus neonatorum cases in a region is data count [14] and is one of the major causes of death in newborns as well as neonatal asphyxia and BBLR. Therefore, it needs to be studied

factors affecting factor the number of cases of neonatal tetanus using generalized Poisson regression models (GP) and Zero-Inflated Poisson Generalized (ZIGP) [15].

METHODOLOGY

This type of research is research of Non-Reactive (unobstructive) [16] for using secondary data such as documents from Health Profile Province in 2015 [5]. The study was conducted using secondary data, along with the number of cases of neonatal tetanus risk factors. The independent variable in this study percentage of visits of pregnant women K4 (X1), the percentage of status TT2 + pregnant women (X2), the percentage of deliveries shaman (X3), the percentage of normal deliveries in health care facilities (X4) and the percentage of neonatal visit (X5). The dependent variable in this study was the number of cases of tetanus neonatorum (Y) [2][17], with the following conceptual framework and analysis steps.
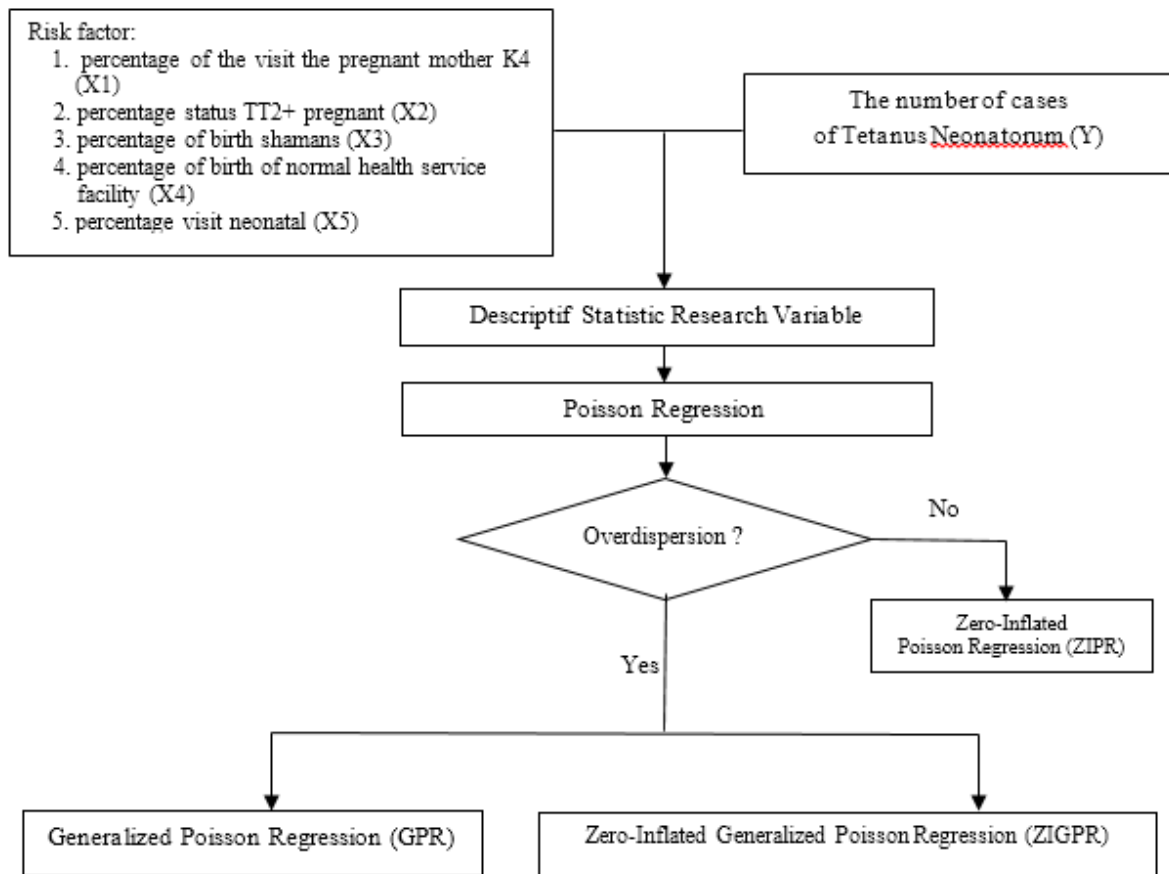

Figure 1. The Framework of operational modeling in the number of cases of Tetanus Neonatorum

Figure 1 can be described as follows:
1. Do descriptive analysis on the research variables
2. Identify the distribution of Poisson on the response variable data used by using Kolmogorov-Smirnov.
3. If data berdistribusi Poisson continued with Poisson regression analysis.
4. Identify the overdispersion using Deviance value.
5. If overdispersion, do GPR an ZIGP analysis using the program R.
   a. Assessing the parameter models with Maximum Likelihood Estimation method
   b. Test the suitability of the model
   c. Test the hypothesis regression model simultaneously and individual

RESULTS AND DISCUSSION

The response variable used in this study is the number of tetanus neonatorum cases in East Java Province in 2015. The number of tetanus neonatorum cases follows the Poisson distribution, as indicated by Kolmogorov Smirnov test with asymptotic value. Sig. (2 tailed) = 0.329 larger than α = 5% then Failed to reject Ho, which means the data follows the Poisson distribution.

The highest number of tetanus neonatorum cases occurred in Bangkalan city with 8 cases, while tetanus neonatorum case was 0 cases in 27 districts / cities. According [10] the data is said to have a value of excess zero if the proportion of data to zero in the response variable more than 50 percent. In detail presented as follows:

Table 1. Zero-Inflation Examination Results on the number of cases of Tetanus Neonatorum

Number of cases of tetanus neonatorum	Frequency	Percentage	Cumulative percentage
0	27	71.1	71.1
1	8	21.1	92.1
2	1	2.6	94.7
4	1	2.6	97.4
8	1	2.6	100.0

Based on Table 2 that the number of observations is 0 on the number of cases of tetanus neonatarum in East Java province in 2015 a total of 27 observation or 71%, resulting in the occurrence of zero-inflation on the response variable.

Modeling the case of neonatarum tetanus in East Java province in 2015 using five predictor variables is the percentage of visits of pregnant women K4 (X_1), the percentage of status TT2 + pregnant women (X_2), the percentage of deliveries shaman (X_3), the percentage of normal deliveries in health facilities (X_4), and the percentage of neonatal visits (X_5). Descriptive statistics of the research variables are presented below.

Table 2. Research Variable descriptive statistics

The variables	N	Min	Max	Mean	Std. Deviation	Variance
The number of cases of tetanus neonatorum (Y)	38	0	8	0.58	1.46	2.14
The percentage of the visit the pregnant mother K4 (X_1)	38	81.26	98.41	90.45	4.30	18.49
The percentage of immunization status TT2+ pregnant (X2)	38	4.00	167.86	74.39	39.73	1578.72
The percentage of birth shamans (X_3)	38	0.00	6.90	0.81	1.67	2.79
The percentage of birth of normal health service facility (X_4)	38	84.01	100.46	94.02	3.84	14.76
The percentage of the visit neonatal (X_5)	38	90.51	106.02	97.29	3.75	14.09

Table 2 shows that in 2015 the average number of cases of tetanus neonatorum is 0.58 percent while variansnya is 2.14 percent. The situation where varians is greater than the mean on the response variable which follow the Poisson distribution can also be one of the causes of overdispersi. But to prove the existence of overdispersi dispersion testing needs to be done using the value of the Deviance divided by degrees free. Deviance value greater than 1 indicates a overdispersi. The following test results using the value of the *Deviance overdispersi*.

Table 3. Test Deviance Overdispersion

The statistics	Value		
Deviance	42.556	32	1.330
Pearson Chi-Square	66.657	32	2.083

Table 3 shows that the value of the Deviance divided by degrees flight is of 1.330. This shows that happen overdispersi. Testing overdispersi served in the following table.

Table 3. Testing Dispersion Parameters

Overdispersion Test	The Value
Z	0.9087
P-value	0.1818
Ω	2.0067

The value of Ω in table 3 is the dispersion parameters where the obtained value is 2.0067 or larger than 0 so that indicates a overdispersi on Poisson regression. Then the modeling with generalized poison regression and ZIGP served in the following table.

Table 6 Results of the parameter estimation of *Poisson Regression, Generalized Poisson, ZIGP Model*

Coefficient Regression (P-value)	Regression		
	Poisson	Generalized Poisson	ZIGP
Count Model			
In Intercept	-17.954 (0.012)	-16.679 (0.064)	-21.746 (0.008)
The percentage of the visit the pregnant mother K4 (X_1)	-0.095 (0.267)	-0.177 (0.156)	-0.243 (0.063)
The percentage of immunization status TT2+ pregnant (X2)	-0.011 (0.182)	0.001 (0.962)	-0.042 (0.000)
The percentage of birth shamans (X_3)	0.443 (0.000)	0.335 (0.033)	0.672 (0.000)
The percentage of birth normal difasilitas health services (X4)	0.137 (0.053)	0.137 (0.142)	0.032 (0.683)
The percentage of the visit neonatal (X_5)	0.134 (0.138)	0.188 (0.131)	-0.012 (0.359)
Zero-Inflated Model			
In Intercept			-442.564 (0.184)
The percentage of the visit the pregnant mother K4 (X_1)			36.235 (0.286)
The percentage of the status of TT2+ pregnant (X_2)			-3.724 (0.292)
The percentage of birth shamans (X_3)			27.769 (0.299)
The percentage of birth of normal health service facility (X_4)			-25.119 (0.318)
The percentage of the visit neonatal (X_5)			-3.398 (0.359)
Pi		1.485	
Omega		0.000	
AIC	80.373	73,000	64,000

Table 6 shows that the factors that affect the cases of tetanus neonatorum in East Java on data that contains many zero value, *ZIGP regression model* provides better results (AIC = 24) if compared with using poisson regression (AIC = 73) and generalized Poisson (AIC = 80.73). The number of cases of tetanus neonatorum using poisson regression influenced by the percentage of the visit the pregnant mother K4 (X1) and the percentage of birth normal health services (X4), using generalized poisson influenced by the percentage of birth shamans (X3). The modeling ZIGP produces two models namely *count model* to predict the data count and *zero-inflated model* to predict the probability of *zero response* on the data. Based on table 6 it is known that examination variables K4 (X1), Status TT2+ Pregnant (X2), and Childbirth Shamans (X3) has the value of the p-value < 0.1 so that it can be said that these variables affect a significant number of cases of tetanus neonatorum.

Next, ZIGP regression model to include the variables significant on the number of cases of tetanus neonatorum is presented in the following table.

Table 7. The results of the Parameter Estimation Zero-Inflated Generalized Poisson Regression Model

The variables	estimate	standard Error	Z value	P-value
Count the model				
In Intercept	-20.642	6,994	-2.952	0.003
The percentage of the visit the pregnant mother K4 (X_1)	-0.252	0.078	-3.215	0.001
The percentage of immunization status TT2+ pregnant (X2)	-0.042	0.009	-4.217	0.000
The percentage of birth shamans (X_3)	0.656	0.146	4.484	0.000
Model Zero-Inflated				
In Intercept	-43.989	32.092	-1.371	0.170
The percentage of the visit the pregnant mother K4 (X_1)	0.534	0.369	1.444	0.148
The percentage of immunization status TT2+ pregnant (X2)	-0.084	0.051	-1.659	0.097
The percentage of birth shamans (X_3)	-0.036	0.827	-0.044	0.965
Log-likelihood	-25.84			

Based on table 7 it is known that a variable percentage of the visit the pregnant mother K4 (X1), the percentage of immunization status TT2+ pregnant (X2) and the percentage of birth shamans (X3) influence significantly to the regression model to count the model, while on the zero-inflated only model variables the percentage of immunization status TT2+ pregnant (X2) is significant.

$$\log(\hat{\mu}_i) = -20.642 - 0.252X_1 - 0.042X_2 + 0.656X_3$$

$$\text{Then } \hat{\mu}_i = \exp(-20.642 - 0.252X_1 - 0.042X_2 + 0.656X_3)$$

$$\log(\hat{p}_i) = -43.989 + 0.534X_1 - 0.084X_2 - 0.036X_3$$

Then $\hat{p}_i = \exp(-43.989 + 0.534X_1 - 0.084X_2 - 0.036X_3)$

The interpretation of the model for data count, every there is increasing the percentage of the visit the pregnant mother K4 (X1) then the average number of cases of tetanus neonatorum will fell 0.777, and every there is increasing the percentage of immunization status TT2+ pregnant (X2) then the average number of cases of tetanus neonatorum will fell 0.958, and each there is increasing the percentage of birth shamans (X3) then the average number of cases of tetanus neonatorum will rise by 1.928. While on the zero-inflated only model variables the percentage of immunization status TT2+ pregnant (X2), which can be interpreted that every there is increasing the percentage of immunization status TT2+ pregnant (X2) then the average number of cases of tetanus neonatorum will fell 0.919.

CONCLUSION

The results of the study showed that the number of cases of tetanus neonatorum is count data and follow Poisson distribution. The number of cases of tetanus neonatorum statistics provide varians greater than mean, deviance value greater than 1 indicates an over dispersion. The case of tetanus neonatorum is the case with many observations worth 0 (zero) i.e. 71%, then Zero-Inflated Generalized Poisson regression model (ZIGP) able to control the value of zero and overdispersion compared with poisson regression and generalized poisson. Model ZIGP regression produces two models, model for data count, every there is increasing the percentage of the visit the pregnant mother K4 (X1) then the average number of cases of tetanus neonatorum will fell 0.777, and every there is increasing the percentage of immunization status TT2+ pregnant (X2) then the average number of cases of tetanus neonatorum will fell 0.958, and each there is increasing the percentage of birth shamans (X3) then the average number of cases of tetanus neonatorum will rise by 1,928. While on the zero-inflated only model variables the percentage of immunization status TT2+ pregnant (X2), which can be interpreted that every there is increasing the percentage of immunization status TT2+ pregnant (X2) then the average number of cases of tetanus neonatorum will fell 0.919.

REFERENCES

[1] Dina, R.A. (2009). Gambaran Epidemiologi Kasus kematian tetanus Neonatorum Di Kabupaten Serang Tahun 2005-2008, Fakultas Kesehatan Masyarakat Universitas Indonesia.

[2] Wibowo, T. (2012). Tetanus Neonatorum. *Buletin Jendela dan Informasi Kesehatan: Eliminasi Tetanus Maternal dan Neonatal Volume 1*. Jakarta: Pusat Data dan Informasi Kementerian RI.

[3] Kementerian Kesehatan Republik Indonesia. (2014). *Profil Kesehatan Indonesia 2014.* Jakarta: Kementerian Kesehatan RI.

[4] Sutjipto, S. (1992). Faktor-Faktor Risiko Yang Berhubungan Dengan Kejadian Tetanus Noenatorum (Doctoral dissertation, Diponegoro University). Diakses dari http://eprints.undip.ac.id.

[5] Dinas Kesehatan Provinsi Jawa Timur. (2016). Profil Kesehatan Provisnsi Jawa Timur Tahun 2015. Surabaya: Dinkes Jawa Timur.

[6] Jansakul, N. dan Hinde, J.P. (2002). Score Tests for Zero-Inflated Poisson Models. Computational Statistics & Data Analysis. *Elsevier Science, 40,* p.75-96.

[7] Triyanto, Purhadi, Bambang Widjanarko Otok, Santi Wulan Purnami. (2015). Parameter estimation of geographically weigthed multivariate Poisson regression. Applied Mathematical Sciences, Vol. 9, no. 81 – 84.

[8] Famoye, F. and Ozmen, I. (2006). Count Regression Models with An Application to Zoological Data Containing Structural Zeros. *Journal of Data Sciences. 5,* p. 491-502.

[9] Khoshgooftar, T.M., Gao.K., dan Szabo, R.M. (2004). Comparing Software Fault Prediction Of Pure and Zero Inflated Poisson Regression Models. *International Journal Of system Science. 36 (11),* p.705-715

[10] Lambert, D. (1992). Zero Inflated Poisson Regression, with an Apllication to Defect in Manufacturing. *Journal Technometric, Feb Vol 34 (1),* p. 1-14.

[11] Famoye, F. and Singh, K.P. (2006). Zero-Inflated Generalized Poisson Regression Model with an Application to Domestic Violence Data. *Journal of data Science. 4,* p. 117-130.

[12] Kartiningrum, E.D. (2012). Pemodelan Faktor Yang Mempengaruhi Kematian Ibu Di Propinsi Jawa Timur Menggunakan Zero Inflated Poisson Regression, *tesis*. FKM Unair, Program Studi Ilmu Kesehatan Masyarakat.

[13] Bozdogan, H. (2000). Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology,* 44, p. 62-91.

[14] Cameron, A.C, & Tivedi, P.K (1998). *Regression Analysis of Count data.* New York: Cambridge University Press.

[15] Bohning, D., Dietz, E., Schlattmann, P. (2012). *Zero Inflated Count Model and Their Applications in Public Health and Social Science*. Paper dalam http://www.ipn.uni-kiel.de

[16] Kuntoro. (2015). Metode Sampling dan Penentuan Besar Sampel. Surabaya: Pustaka Melati.

[17] Abrutyn, E. (2008). Tetanus. *In Fauci, A.S., et al ed. Harrison's Principles of Internal Medicine. 17th ed*. America: McGrawHill, 898-899.