



# Generalized Linear Models of Malaria Incidence in Jubek State, South Sudan

**Loro Gore Lado Jumi**

Department of Statistics and Demography, University of Juba, Juba, South Sudan

**Email address:**

lorogore@gmail.com

**To cite this article:**

Loro Gore Lado Jumi. Generalized Linear Models of Malaria Incidence in Jubek State, South Sudan. *Science Journal of Applied Mathematics and Statistics*. Vol. 5, No. 4, 2017, pp. 134-138. doi: 10.11648/j.sjams.20170504.12

**Received:** May 8, 2017; **Accepted:** May 20, 2017; **Published:** July 7, 2017

---

**Abstract:** Malaria is a leading cause of morbidity and mortality in South Sudan. This study is meant to focus on the trend of malaria incidence in Jubek state, South Sudan. Data on weekly malaria incidence for the period January 2011 to October 2015 were used in the study. Generalized linear models, Poisson and negative binomial regression models were employed to analyze the data. Results obtained suggest that malaria incidence has been still on increase by 0.0030 and 0.0032 per week respectively. Additionally, incidence rate ratio suggests an increase of 0.3% per week of malaria incidence in Jubek state. The study recommends malaria control programmes focused on reducing malaria incidence be introduced in South Sudan in general and in Jubek state in particular.

**Keywords:** Malaria Incidence, Poisson Regression, Negative Binomial Regression, South Sudan

---

## 1. Introduction

Malaria incidence is the number of new malaria cases diagnosed during a given time interval in relation to the unit of population in which they occur [3]. Globally, an estimated 3.2 billion people are at risk of being infected with malaria and developing the disease, with 1.2 billion at high risk (greater than a one-in-thousand chance of getting malaria in a year). According to the latest estimates, 198 million cases of malaria occurred in 2013 worldwide and the disease led to 584,000 deaths [15]. The WHO African region bears the highest burden of malaria, accounting for an estimated 90% of all malaria deaths worldwide. More than three quarters (78%) of malaria deaths occurred among children aged less than five years [15]. The estimated total number of malaria cases fell from 227 million in 2000 to 198 million in 2013. During the same period, the population at risk of malaria infection increased by 25% globally and by 43% in the WHO African region [15]. The incidence of malaria which takes into account population growth is estimated to have decreased by 37% between 2000 and 2015. As a result, malaria is no longer the leading cause of death among children in sub-Saharan Africa [14]. However, despite the decrease in the malaria cases in sub-Saharan Africa, still malaria is the leading cause of morbidity and mortality in

South Sudan, accounting for 20% to 40% morbidity with over 20% deaths and 30% of all hospital admissions. The disease is endemic country-wide putting the entire population at risk of infection and exacting a greater toll in children under five years and pregnant women. Parasite prevalence ranges from less than 1% to more than 40% with great variability across the states and is higher in rural areas than in urban areas. *Plasmodium falciparum* is the most dominant and is responsible for up to 94% of all morbidity [9].

Several studies have been carried out on malaria incidence. [8] investigated the effects of climate on malaria in Burundi using generalized linear models and generalized additive mixed models. The results suggest a strong positive association between malaria incidence in a given month and minimum temperature of the previous month. In contrast, it was found that rainfall and maximum temperature in a given month have possible negative effect on malaria incidence of the same month. [10] studied the trend of malaria prevalence in Minna, Nigeria, by employing Poisson and Negative binomial regression models. The results revealed that the prevalence of malaria is still on increase by 6% on monthly basis. [6] examined the relationship between malaria and environmental and socio-economic variables in the Sudan using health production modified model. The regression results showed significant relationships between malaria and rainfall and

water bodies. Other variables including Human Development Index, temperature, population density and percent of cultivated areas were not significant. [11] developed polynomial regression model to trace the impact of environmental factors on incidence of malaria cases in Lakhimpur district of Assam in India. The study revealed that variables such as minimum temperature, maximum temperature, average rainfall and population play an important role in predicting malaria incidence cases. [5] carried out a study to identify the spatial and trends of malaria incidence in Nepal Poisson and negative binomial regression models were used to fit malaria incidence rates as a function of year and location. The study showed a steady decreasing trend in malaria incidence, but the numbers of cases are still very high. [13] studied malaria incidence over time and its association with temperature and rainfall in four counties of Yunnan province, China. Seasonal trend decomposition was used to examine secular trends and seasonal patterns in malaria incidence, a Poisson regression with Distributed lag non-linear models were used to estimate the weather drivers of malaria seasonality. The study revealed that there was a declining trend in malaria incidence in all four counties. [1] conducted a study on malaria vector control in South Sudan. The study revealed that the peak of malaria transmission season lasting 7 to 8 months of the year south of the country and 5 to 6 months in the north. The main vectors of malaria are *Anopheles gambiae* s.s., *Anopheles arabiensis* and *Anopheles funestus* s.s. [2] carried out a study using logistic regression to estimate and assess malaria prevalence and the use of malaria risk reduction measures and their association with selected background characteristics in South Sudan. The results suggest that educational attainment need not be very advanced to affect practices of malaria prevention and treatment. Primary school attendance was a stronger predictor for use of malaria risk reduction measures than any other selected background characteristics. South Sudan being a new country, few studies on malaria incidence have been carried out. However, no specific study has ever been carried out in Jubek State. Therefore, the study undertaken is meant to focus on the trend of malaria incidence in Jubek State using generalized linear models. Determination of malaria trend is critical for the design of policies aimed at reducing malaria incidence.

## 2. The Methodology of the Study

In this study, data on weekly malaria incidence in Jubek state for the period January 2011 to October 2015 were obtained from the Directorate of Preventive Health Services of Ministry of Health, South Sudan. The response variable is the malaria incidence, and the explanatory variable is time in weeks coded from 0 to 252 weeks. The generalized linear models were applied. The basic count data regression models can be represented and understood using generalized linear models (GLM) framework [7]. Poisson regression is commonly used for modeling the number of cases of disease in a specific population with a certain time. Poisson regression is a special case of (GLM) where the response

variable follows Poisson distribution. Poisson models for disease counts are often over-dispersed, in which case the negative binomial regression is more appropriate [12]. The negative binomial model is an extension of Poisson model for incidence rates that allows for the over-dispersion that commonly occurs for disease count.

Poisson probability distribution is specifically suited to count data, with density function

$$f(Y_i) = \frac{\mu^Y e^{-\mu}}{Y!}, Y = 0, 1, 2, \dots \quad (1)$$

Where  $f(Y)$  denotes the probability that the discrete random variable  $Y$  takes non-negative integer values,  $\mu$  is the parameter of Poisson distribution. It can be proved that

$$E(Y) = \mu \quad (2)$$

$$var(Y) = \mu \quad (3)$$

A unique feature of Poisson distribution is that the mean and the variance are the same [4].

In the Poisson regression model, the number of events  $y$  has a Poisson distribution with conditional mean that depends on an individual's characteristics:

$$\mu_i = E(y_i | x_i) = \exp(x_i \beta) \quad (4)$$

It can also be written as

$$\ln(\mu_i) = x_i \beta \quad (5)$$

Under this model as  $\mu_i$  increases, the conditional variance of  $y$  increases. The Poisson regression model can be thought of as a non-linear model [16].

The Negative binomial regression model allows the conditional variance of  $y$  to exceed the conditional mean. The mean  $\mu$  is replaced with the random variable  $\tilde{\mu}$ :

$$\tilde{\mu}_i = \exp(x_i \beta + \varepsilon_i) \quad (6)$$

Where  $\varepsilon$  is a random error that is assumed to be uncorrelated with  $x$ .

$$\tilde{\mu}_i = \exp(x_i \beta) \exp(\varepsilon_i) = \mu_i \exp(\varepsilon_i) = \mu \delta_i \quad (7)$$

In the Negative binomial regression model the most convenient assumption is that the mean of the error term is 1. Hence

$$\tilde{\mu}_i = \exp(x_i \beta) \exp(\varepsilon_i) = \mu_i \exp(\varepsilon_i) = \mu \delta_i = \mu_i \quad (8)$$

The assumption is that  $\delta$  has a gamma distribution with parameter.  $(\delta) = 1$  and  $Var(\delta) = \frac{1}{v}$ .

The expected value of  $y$  for the Negative Binomial distribution is the same as for Poisson distribution, but the conditional variance differs:

$$Var(y_i | x_i) = \mu_i \left( 1 + \frac{\mu_i}{v_i} \right) = \exp(x_i \beta) \left( 1 + \frac{\exp(x_i \beta)}{v_i} \right) \quad (9)$$

Since  $\mu$  and  $v$  are positive, the conditional variance of  $y$  must exceed the conditional mean,  $v$  is the same for all individuals:

$$v_i = \alpha^{-1} \text{ for } \alpha > 0 \quad (10) \quad \alpha \text{ is known as the dispersion parameter since increasing } \alpha \text{ increases the conditional variance of } y.$$

$$Var(y_i/x_i) = \mu_i \left(1 + \frac{\mu_i}{\alpha^{-1}}\right) = \exp(x_i\beta) \left(1 + \frac{\exp(x_i\beta)}{v_i}\right) = \mu_i(1 + \alpha\mu_i) = \mu_i + \alpha\mu_i^2 \quad (11)$$

If  $\alpha = 0$  then the mean and variance are the same [16].

### 3. Results and Discussion

In this study, weekly data on malaria incidence rate in Jubek state, South Sudan were analyzed using STATA 10.0. Poisson and Negative binomial regression models were employed to analyze malaria incidence rate with respect to

time. In table 1 the time variable showed significance with a coefficient of .003, (p-value=.000 <0.05), this suggest that time has positive effect on malaria incidence. It is clear that for a unit increase in time (Week) the mean (or rate) of malaria incidence increases by .003. The estimated Poisson model is

$$\ln(\text{Malaria incid}) = 5.359 + .003x$$

**Table 1.** Results of Poisson regression model fitted to malaria incidence rate.

| Poisson mir time           |  |  |                            |  |  |  |
|----------------------------|--|--|----------------------------|--|--|--|
| Iteration 0:               |  |  | log likelihood = -5726.393 |  |  |  |
| Iteration 1:               |  |  | log likelihood = -5726.393 |  |  |  |
| Poisson regression         |  |  | Number of obs = 248        |  |  |  |
|                            |  |  | LR chi2 (1) = 3776.86      |  |  |  |
|                            |  |  | Prob > chi = 0.0000        |  |  |  |
| Log likelihood = -5726.393 |  |  | Pseudo R2 = 0.2480         |  |  |  |

  

| mir  | Coef.     | Std. Err. | z      | P >  z | [95% Conf. | Interval] |
|------|-----------|-----------|--------|--------|------------|-----------|
| time | 0.0030723 | 0.0000504 | 61.02  | 0.000  | 0.0029736  | 0.003171  |
| cons | 5.359782  | 0.0078985 | 678.58 | 0.000  | 5.344301   | 5.375263  |

  

| Poisgof                         |  |  |  |  |  |  |
|---------------------------------|--|--|--|--|--|--|
| Goodness-of-fit chi2 = 9590.286 |  |  |  |  |  |  |
| Prob > chi2 (246) = 0.0000      |  |  |  |  |  |  |
| Poisgof, pearson                |  |  |  |  |  |  |
| Goodness-of-fit chi2 = 9028.633 |  |  |  |  |  |  |
| Prob > chi2 (246) = 0.0000      |  |  |  |  |  |  |

**Table 2.** Results of Negative Binomial regression model fitted to malaria incidence rate.

| nbreg mir time               |  |  |                             |  |  |  |
|------------------------------|--|--|-----------------------------|--|--|--|
| Fitting Poisson model:       |  |  |                             |  |  |  |
| Iteration 0:                 |  |  | log likelihood = -5726.393  |  |  |  |
| Iteration 1:                 |  |  | log likelihood = -5726.393  |  |  |  |
| Fitting constant-only model: |  |  |                             |  |  |  |
| Iteration 0:                 |  |  | log likelihood = -1678.4435 |  |  |  |
| Iteration 1:                 |  |  | log likelihood = -1559.1457 |  |  |  |
| Iteration 2:                 |  |  | log likelihood = -1555.816  |  |  |  |
| Iteration 3:                 |  |  | log likelihood = -1555.7729 |  |  |  |
| Iteration 4:                 |  |  | log likelihood = -1555.7729 |  |  |  |
| Fitting full model:          |  |  |                             |  |  |  |
| Iteration 0:                 |  |  | log likelihood = -1522.7537 |  |  |  |
| Iteration 1:                 |  |  | log likelihood = -1517.3108 |  |  |  |
| Iteration 2:                 |  |  | log likelihood = -1517.0811 |  |  |  |
| Iteration 3:                 |  |  | log likelihood = -1517.0808 |  |  |  |
| Iteration 4:                 |  |  | log likelihood = -1517.0808 |  |  |  |
| Negative binomial regression |  |  | Number of obs = 248         |  |  |  |
|                              |  |  | LR chi2 (1) = 77.38         |  |  |  |
| Dispersion = mean            |  |  | Prob > chi = 0.0000         |  |  |  |
| Log likelihood = -1517.0808  |  |  | Pseudo R2 = 0.0249          |  |  |  |

| mir      | Coef.     | Std. Err. | z      | P >  z | [95% Conf. | Interval] |
|----------|-----------|-----------|--------|--------|------------|-----------|
| time     | 0.0031536 | 0.0003323 | 9.49   | 0.000  | 0.0025023  | 0.0038049 |
| cons     | 5.349097  | 0.047622  | 112.32 | 0.000  | 5.255759   | 5.442434  |
| ln alpha | -2.023918 | 0.0907079 |        |        | -2.201702  | -1.846134 |
| alpha    | 0.1321367 | 0.0119858 |        |        | 0.1106147  | 0.1578462 |

Likelihood- ratio test of alpha = 0: chibar2 (01) = 8418.62 Prob>=chibar2 = 0.000

**Table 3.** Results of Negative binomial regression model for incidence rate ratio of malaria.

| nbreg mir time, irr          |  |  |  |                             |  |  |
|------------------------------|--|--|--|-----------------------------|--|--|
| Fitting Poisson model:       |  |  |  |                             |  |  |
| Iteration 0:                 |  |  |  | log likelihood = -5726.393  |  |  |
| Iteration 1:                 |  |  |  | log likelihood = -5726.393  |  |  |
| Fitting constant-only model: |  |  |  |                             |  |  |
| Iteration 0:                 |  |  |  | log likelihood = -1678.4435 |  |  |
| Iteration 1:                 |  |  |  | log likelihood = -1559.1457 |  |  |
| Iteration 2:                 |  |  |  | log likelihood = -1555.816  |  |  |
| Iteration 3:                 |  |  |  | log likelihood = -1555.7729 |  |  |
| Iteration 4:                 |  |  |  | log likelihood = -1555.7729 |  |  |
| Fitting full model:          |  |  |  |                             |  |  |
| Iteration 0:                 |  |  |  | log likelihood = -1522.7537 |  |  |
| Iteration 1:                 |  |  |  | log likelihood = -1517.3108 |  |  |
| Iteration 2:                 |  |  |  | log likelihood = -1517.0811 |  |  |
| Iteration 3:                 |  |  |  | log likelihood = -1517.0808 |  |  |
| Iteration 4:                 |  |  |  | log likelihood = -1517.0808 |  |  |
| Negative binomial regression |  |  |  | Number of obs = 248         |  |  |
|                              |  |  |  | LR chi2 (1) = 77.38         |  |  |
|                              |  |  |  | Prob > chi2 = 0.0000        |  |  |
|                              |  |  |  | Pseudo R2 = 0.0249          |  |  |
| Dispersion= mean             |  |  |  |                             |  |  |
| Log likelihood = -1517.0808  |  |  |  |                             |  |  |

  

| mir      | IRR       | Std. Err. | z    | P >  z | [95% Conf. Interval] |           |
|----------|-----------|-----------|------|--------|----------------------|-----------|
| time     | 1.003159  | 0.0003333 | 9.49 | 0.000  | 1.002505             | 1.003812  |
| ln alpha | -2.023918 | 0.0907079 |      |        | -2.201702            | -1.846134 |
| alpha    | 0.1321367 | 0.0119858 |      |        | 0.1106147            | 0.1578462 |

Likelihood- ratio test of alpha = 0: chibar2 (01) = 8418.62 Prob > = chibar2 = 0.000

In table 2 the explanatory variable is significant with a coefficient of.0032 ( $p - value = .000 < .05$ ), this suggests that time has positive effect on malaria incidence. It is evident that for a unit increase in time (Week) the mean (or rate) of malaria incidence increases by.0032 units. The estimated Negative binomial regression model is

$$\ln(\text{malaria incid.}) = 5.349 + .0032x - 2.0239\ln\alpha$$

Poisson regression and Negative binomial models suggest that malaria incidence has increased by 0.003 and 0.0032 per week, respectively. These results may be attributed to factors such as hot and humid climate, high precipitation and socio-economic status in Jubeck state. The incidence rate ratio results in Table 3 suggest that malaria incidence increases by 0.3% with every unit increase in time (week). That is, malaria incidence in Jubeck state increased by 0.3% every week. The results of this study show consistency with the findings of the study in Nigeria [6].

## 4. Conclusion

The objective of this study was to study the trend of malaria incidence in Jubeck state, South Sudan. This study has used weekly data on malaria incidence for the period January 2011 to October 2015 in Jubeck state, South Sudan. The regression is based on generalized linear models Poisson and negative binomial regression. The results obtained suggest that malaria incidence is increasing with time. South Sudan gained its independence in 2011 from Sudan. The country healthcare system has faced a lot of challenges such as lack of infrastructure, inadequate equipment, human resources and weak malaria vector control. Therefore, the increasing

malaria trend may be attributed to poor health services, socio-economic and environmental factors. Determination of malaria trend is critical for the design of policies aimed at reducing malaria incidence and improvement of healthcare system. There is need for study that incorporates more explanatory variables.

## Acknowledgments

The author is grateful to Dr. Thomas Hakim of the Directorate of Preventive Health Services, Ministry of Health, South Sudan, for providing data on malaria cases used in the study, and Mr. Emmanuel Pitia Zacharia Lado, Lecturer in Economics, University of Juba, for carrying out data analysis.

## References

- [1] Chanda, E., Daggale, C., Pasquale, H., Aairwe, R., Baba, S and Mnzava, A.2013. Addressing malaria vector control challenges in South Sudan; proposed recommendations. Malaria Journal 12: 59.
- [2] Draebel T, Kueil, B., G. and Meyrowitsch D. W. (2013) Prevalence of Malaria and use of Malaria Risk Reduction Measures among resettled pregnant woman in South Sudan.
- [3] Gosoni, L. 2008. Development of Bayesian geostatistical models with applications in malaria epidemiology. Doctoral Dissertation, Swiss Tropical Institute, University of Basel.
- [4] Gujarati, D. N. (2004) Basic Econometrics, 4<sup>th</sup> edition, McGraw-Hill Companies.

- [5] Kakchapati, S. & Ardkaew, J. 2011. Modeling of Malaria Incidence in Nepal. *Journal of Research in Health Sciences* 11 (1): 7-13.
- [6] Musa M, I, Shohaimi S., Hashim N, R., and Krishnarajah I (2012) Environmental and Socio-Economic Determinants of Malaria Rate in Sudan. *Research Journal of Environmental and Earth Sciences* 4 (11): 923-929.
- [7] Nelder, J., A. et al. 1992. Generalized linear models. In *Breakthrough in Statistics*: 547-563, Springer New York.
- [8] Nkurunziza H., Gebhardt A., Pilz J. 2010. Bayesian modelling of the effect of climate on malaria in Burundi. *Malaria Journal* 9: 114.
- [9] Pasquale H., Jarvese M., Julla A., Doggale C., Sebit B., Lual M. Y., Baba S. P. and Chand E. 2013. Malaria control in South Sudan, 2006-2013: Strategies, progress and challenges. *Malaria Journal* 12: 374.
- [10] Patience, E. O. and Osagie A. M. 2014. Modeling the prevalence of malaria in Niger State: An application of Poisson regression and negative binomial regression models. *International Journal of Physical Sciences* 2 (4): 061-068.
- [11] Saikia N. J., Hazarika J., Das P. K., and Hussain S. 2015. Role of environmental factors on incidence of malaria cases: a case study using polynomial regression. *International Journal of Current Research* 7 (12): 24157-24160.
- [12] Vennables, W. N. and Ripley, B. D. 2002. *Modern Applied Statistics with S*. 4<sup>th</sup> ed. New York, Springer.
- [13] Wardrop N. A., Barnett A. G., Atkinson J. A., and Clements A. C. 2013. *Plasmodium vivax* malaria incidence over time and its association with temperature and rainfall in four Counties of Yunnan Province, China. *Malaria Journal* 12: 452.
- [14] WHO. 2015. Global malaria programme. World malaria report.
- [15] WHO. 2015. World health statistics.
- [16] Williams, R. (2016) *Models for Count Outcomes*, University of Notre Dame, USA.