

## Cutaneous Leishmaniasis Modeling: the case of Msila Province in Algeria

H. Elhadj<sup>1</sup>, Y. Kerboua Ziari<sup>2</sup>, and S. Selmane<sup>3</sup>

<sup>1</sup>High National School of Statistics and Applied Economics, Algiers, Algeria

<sup>2</sup>Faculty of Physics, University of Sciences and Technology Houari Boumediene, Algiers, Algeria

<sup>3</sup>Faculty of Mathematics, University of Sciences and Technology Houari Boumediene, Algiers, Algeria

---

Copyright © 2015 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT:** Cutaneous leishmaniasis is one of the infectious diseases that affects public health and represents a real threat especially in developing countries. The disease is transmitted by the bite of certain species of sandflies and occurs predominantly in warm, humid and tropical climate.

Finding the source of cutaneous leishmaniasis and identifying factors that promote its spread could help to a good prediction of the epidemic in time. The aim of this study is the construction of a statistical model that reproduces the number of affected cases using climate factors influencing the presence of sandflies.

Given the extensive development of the Generalized Linear Models and their performance in modeling count data as well as their adaptation to the problem of overdispersed data, we present the utility and the basic foundations of Poisson and quasi-Poisson regression models. Thereafter, we build a forecasting model that could predict the number of monthly cases of the cutaneous leishmaniasis from climatic factors during the period 2008-2011 in the province of Msila which is one of the Algerian provinces heavily affected by the epidemic in question. In our case the temperature and trend factor were retained in the model. Poisson regression gave a good result after eliminating the effect of overdispersion.

**KEYWORDS:** Cutaneous leishmaniasis, Count data, Generalized Linear Models, Poisson regression, Overdispersion.

### 1 INTRODUCTION

Vector-borne diseases are illnesses caused by pathogens and parasites in human populations transmitted by "vectors" (insects) from an infected person (or an infected animal) to another. More than a billion cases affected by these diseases are detected annually with more than one million recorded worldwide deaths [7].

Cutaneous leishmaniasis ranges over tropical and warm temperate countries, it is reported in 88 countries in 5 sources: Mediterranean, Chinese, Indian, Central and South American, and African. Its prevalence is estimated at 12 million with an incidence of 2 million (1.5 million of cutaneous leishmaniasis with 90% in Algeria, Afghanistan, Saudi Arabia, Brazil, Iran, Peru, Syria and 500,000 visceral leishmaniasis with 90% in Bangladesh, Brazil, India, Nepal, Sudan) [7].

In Algeria, the vector-borne diseases continue to grow, the incidence was multiplied by five between 1989 and 2002, from 5.23 to 27.12 cases per 100 000 inhabitants. These diseases include leishmaniasis that occurs in three clinical forms, human leishmaniasis (cutaneous and visceral leishmaniasis) and canine leishmaniasis [2].

Cutaneous leishmaniasis is transmitted through the bites of infected female phlebotomine sandflies that are most active in humid environment during the warmer months and at night. The adults, who appear only at about May, are active during warmer months then disappear in autumn. In 1992, *Phlebotomus papatasi* has been identified as the main vector of the disease in Algeria, it causes an exclusive skin damage it is found in the steppe regions, arid and semi-arid, mainly at the northern fringe of the Sahara and admits as reservoir wild rodents. The incubation period of the disease takes from two to six months [2], [6].

Three cutaneous leishmaniasis outbreaks occurred between 2004 and 2006, with respectively 14822, 25511, and 14714 cases. Cited after former sources, Biskra at the east and Ababla at the west, Msila experienced the epidemic in 1982 with 8000 recorded cases. The disease used to be mainly endemic in the sub-Saharan steppe, however, a geographical spread towards the north and west has taken place recently [10].

## 2 METHODOLOGY

### 2.1 STUDY AREA

The area concerned by the study is the province of Msila, one of the main sources for cutaneous leishmaniasis. It is located at the highlands of central-eastern Algeria and extends over a land area of 18,175 km<sup>2</sup>. Its population is estimated at 1,107,821 [9].

Its climate is mostly arid, and partly semi-arid and Saharan. The precipitations are low; its average temperature is about 34°C in summer season which is dry and very hot and 10°C in winter season which is very cold.

### 2.2 DATA

Epidemiological data concerning the number of cases registered by the cutaneous leishmaniasis 'LC' in the period 2008-2011 were obtained from the National Institute of Public Health [9]. The weather data for the same period involving temperature (T), humidity (H), and rainfall (P) were obtained from the National Meteorological Office [8].

### 2.3 STATISTICAL METHOD PROBLEM OF OVERDISPERSION OF DATA

The generalized linear model (GLM) was developed as a way to unify the statistical models including linear regression, logistic regression, log-linear regression, Poisson regression. It proposes an iterative method called least squares method iteratively re-weighted to estimate the maximum likelihood of the model parameters.

The GLM relates response variable Y with a set of explanatory variables X<sub>j</sub> under the following conditions:

- The distribution of Y is a member of an exponential family, such as Normal, Binomial, Poisson, Gamma, or inverse-Gaussian families of distribution.
- The fitted mean of the model  $\mu$  is connected to a linear function of regressors  $\eta$  (linear predictor)

$$\eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$$

- The link function which transforms the expectation of the response variable to the linear predictor is monotonic and differentiable.

$$g(\mu_i) = \eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}.$$

All of GLM distributional families are separate parameterizations of an underlying single parameter exponential probability distribution that is commonly expressed as

$$f(y|\theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right] \quad (1)$$

Where

- $\vartheta$  is the canonical parameter or link function
- $b(\vartheta)$  is the cumulant
- $a(\phi)$  is the scale parameter, set to one in discrete and count models
- $C(y, \phi)$  is the normalization term, guaranteeing that the probability function sums to unity.

We can define the various distributions of the exponential family (normal, Poisson, binomial ... etc) by specifying the functions a, b and c

Some densities, such as the negative binomial distribution and the Weibull distribution, are not members of the exponential family, but they are close to the GLM.

**2.3.1 THE POISSON REGRESSION**

Poisson regression is the basic count model upon which a variety of other count models are based.

The Poisson distribution may be characterized as:

$$P(Y = Y_i | X_i) = \frac{e^{-\mu_i} \mu_i^{Y_i}}{Y_i!}, Y = 0, 1, \dots, i = 1, \dots, N \quad (2)$$

Where  $\mu$  is the mean parameter. The scale parameter is assumed to be 1 for this distribution.

The relationship between, the fitted mean of the model  $\mu$ , parameters  $\beta$ , and model covariates or predictors  $x$ , is parameterized such that

$$\mu = \exp(x\beta)$$

The exponential form of Poisson distribution may be defined as follows:

$$f(y|\theta, \phi) = \exp(y\log\mu - \mu - \log y!) \quad (3)$$

The components of the model concluded from the general form (Equation (1)) are given by:

$$\theta = \log(\mu), \phi \equiv 1, a(\phi) = 1, b(\theta) = \exp(\theta), c(y, \phi) = -\log y!$$

With a mean equal to  $b'(\theta) = \mu$  and a variance  $b''(\theta) = \mu$ .

The parameters  $\beta$  of a GLM can be estimated using maximum likelihood by applying the Newton Raphson method with Fisher scoring; McCullagh and Nelder show that the optimization is equivalent to iteratively reweighted least squares (IRWLS) [5].

The Poisson log-likelihood function may then be derived as

$$L(\mu; y) = \sum_{i=1}^n \exp \{y_i \ln(\mu_i) - \mu_i - \ln(y_i!)\} \quad (4)$$

Since the Poisson has a link defined as  $\ln(\mu)$ , the inverse link is  $\mu = \exp(x\beta)$ , Substituting into equation (4) yields

$$\begin{aligned} L(\beta, y) &= \sum_{i=1}^n (y_i(x'_i\beta) - \exp(x'_i\beta) - \ln(y_i!)) \\ &= \sum_{i=1}^n \{y_i(x'_i\beta) - \exp(x'_i\beta) - \log \Gamma(y_i + 1)\} \end{aligned} \quad (5)$$

Its first derivative is:

$$\frac{\partial(L(\beta, y))}{\partial \beta} = \sum_i^n (y_i - \exp(x'_i\beta))x_i \dots (6)$$

Solving for parameter estimates,  $\beta$ , entails setting the above to zero.

The second derivative is used to estimate standard errors of parameters.

**2.3.2 PROBLEM OF OVERDISPERSION OF DATA**

Equality of the mean and the variance is referred to as the equidispersion property of the Poisson distribution.

Since count data typically has a variance that exceeds the mean this property is frequently violated, this problem could be caused by a positive correlation or excessive variation between observations which affects standard errors of the parameter estimates.

To deal with overdispersion in count response models we can adjust Poisson regression by scaling of standard errors. Another approach uses negative binomial model, which is based on a Poisson distribution with parameter  $\mu$  considered as a random variable gamma  $(\gamma_i, \delta)$  where  $\gamma_i = e^{x_i\beta + \alpha}$  [5].

### 2.3.3 GOODNESS OF FIT

Goodness-of-fit measures for GLMs are the deviance and Pearson statistics. The formula for the deviance statistic is defined as:

$$D = 2 \sum_{i=1}^n \{\mathcal{L}(y_i; y_i) - \mathcal{L}(\mu_i; y_i)\} \quad (7)$$

With  $L(y_i; y_i)$  indicating a log-likelihood function for the saturated model which refers to perfectly fitting model using  $n$  free parameters to fit  $n$  observations with every value of  $\mu$  given the value  $y$  in its place.  $L(\mu_i; y_i)$  is the log-likelihood function for the model being estimated; in this model the log-likelihood function achieves its maximum achievable value.

The formula for Pearson statistics is:

$$P = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\omega}_i} \quad (8)$$

Where  $\hat{\mu}_i$  and  $\hat{\omega}_i$  are estimates of  $\mu_i$  and  $\omega_i$ . In Poisson regression model  $\omega_i = \mu_i$ , so that

$$P_p = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \quad (9)$$

It is compared with  $(n-k)$ , reflecting a degree of freedom correction [2], [5].

## 3 RESULTS

### 3.1 DESCRIPTIVE ANALYSIS OF DATA

An elementary examination of figure 1 allows to conclude that the evolution of the temperature (which favors the activity of sandflies) looks like that of the series representing the number of cases affected by cutaneous leishmaniasis but with a certain interval of time, this corresponds to the necessary duration of incubation of cutaneous leishmaniasis. It should be noted that for year 2009 the average temperature of July approached 36°C (while generally it is around 34°C) and that we observe for the same year a very high number of cases exceeding 2000 cases. The humidity evolves in the inverse direction of temperature evolution.

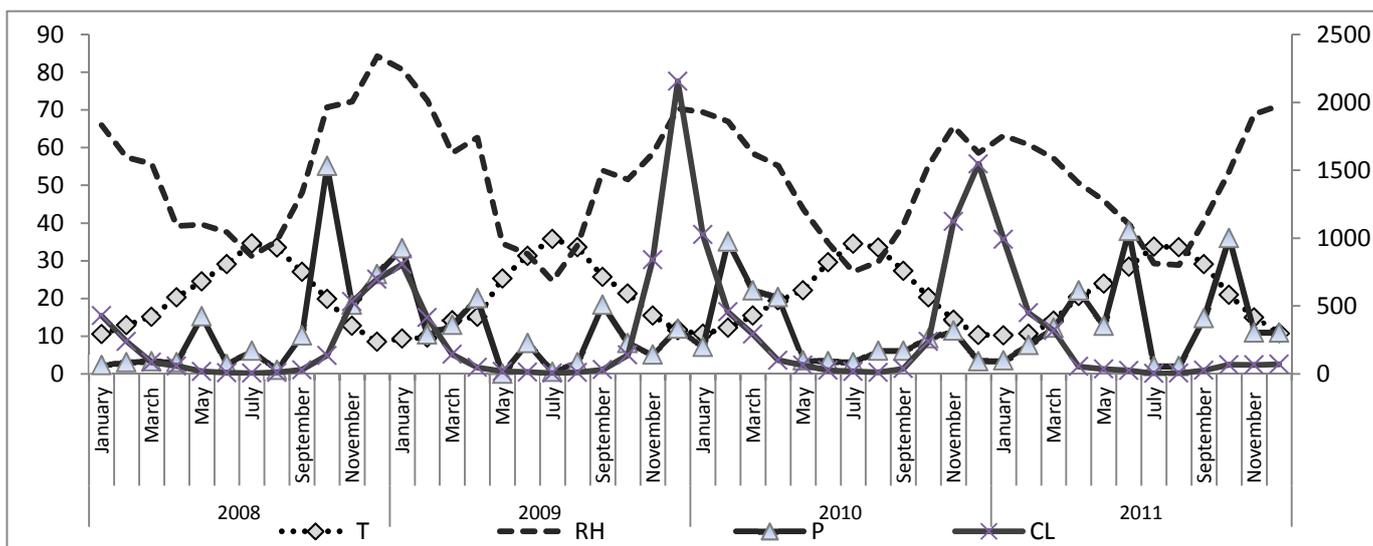


Figure 1. Evolution of the number of cases of cutaneous leishmaniasis and climatic factors (temperature, relative humidity and precipitation) from January 2008 to December 2011.

To examine the time gap between exposure to weather conditions and following occurrence of cutaneous leishmaniasis we have used cross correlation function represented by the correlogram (Figure 2) which computes autocorrelations for data values at varying time lags.

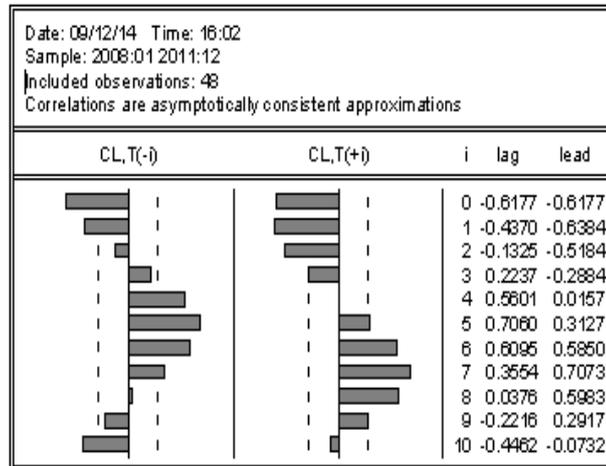


Figure 02. Cross correlogram of cutaneous leishmaniasis and temperature.

Regarding series RH and T, their effect on the LC series have a lag of 5, the cross correlation between CL and  $T_{(-5)}$  is equal to 0.7 confirming that the incubation period of the disease is of 5 months while it takes the value of -0.63 between LC and  $RH_{(-5)}$  which says that humidity has the opposite role of temperature. When the humidity is minimal, the number of cases increases after the incubation period. The cross correlogram between P and CL does not reveal any relationship.

3.1.1 MODEL CONSTRUCTION

After the introduction of all variables into the model, we proceeded through the elimination of variables P and RH that were no significant. In addition to the temperature (T), the constant and trend (t) variable with a quadratic form were taken account in the model.

The first results assert the presence of overdispersion (Pearson chi-square / degree of freedom ratio was far from 1). An adjustment to the Poisson regression by introducing a dispersion parameter (equal to the inverse of Pearson Chi square / degree of freedom ratio) allowed us to eliminate the effect of the overdispersion.

The final model is then written as:

$$\text{Log CL} = 0.2 + 0.17 T_{(-5)} - 0.004 t^2 + 0.16 t \quad (10)$$

The parameters of the Deviance and Pearson Chi-Square (Table 1) have a value around 1, thereby affirming the goodness of the fitted model.

Table 1. Information on the goodness of the fit

	Value	df	Value/df
Deviance	39,245	39	1,006
Scaled Deviance	39,245	39	
Pearson Chi-Square	39,000	39	1,000
Scaled Pearson Chi-Square	39,000	39	
Log Likelihood <sup>a</sup>	-22,114		
Akaike's Information Criterion (AIC)	52,228		
Finite Sample Corrected AIC (AICC)	53,281		
Bayesian Information Criterion (BIC)	59,273		
Consistent AIC (CAIC)	63,273		

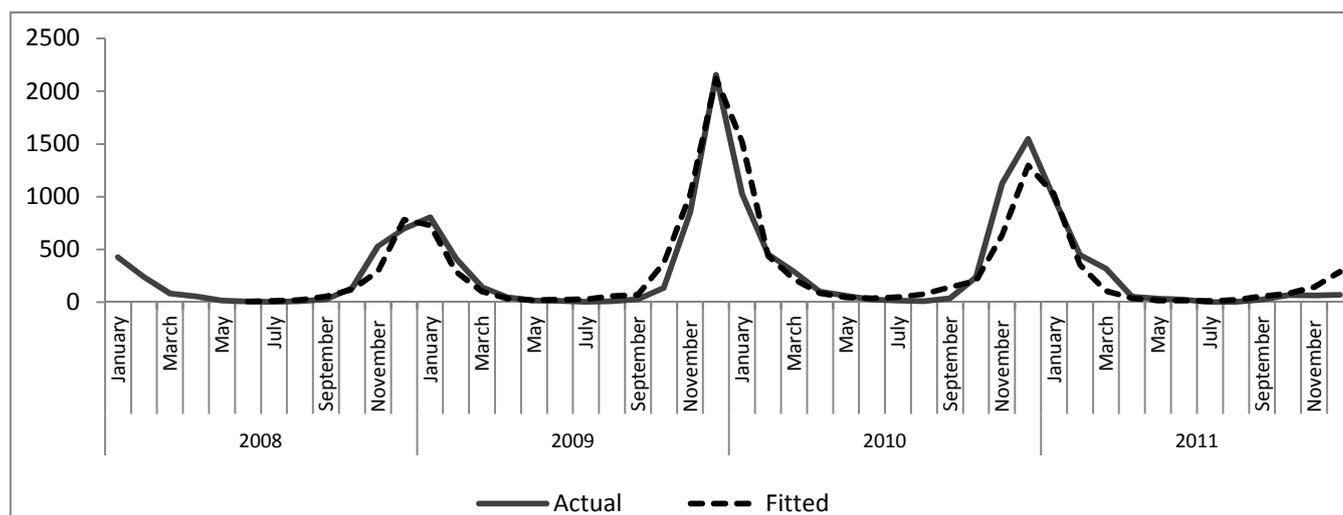


Figure 3. Comparison between the actual values and those estimated by the model from June 2008 to December 2011.

The figure 3 shows the closeness between the actual values of cutaneous leishmaniasis cases and those computed from the adjusted model.

#### 4 CONCLUSION

Cutaneous leishmaniasis remains an epidemic which constitutes a real sanitary problem in some regions of Algeria among which the province of Msila. To look for its fundamental causes, we used a statistical model based on a generalized linear model with a Poisson regression that led us to retain the temperature as the main climatic factor and a trend variable, in quadratic form, to account for non-climatic factors such human behavior, degradation of the environment, and other factors that could influence the number of sandflies bites. To solve the problem of overdispersion present in the model that is responsible for the underestimation of the standard errors of the estimators, we introduced the overdispersion parameter which leads to a quasi Poisson regression. The predicted values are close to the actual values, which confirm the adequacy of the model for the study period.

Equation (10) predicts that an increase in temperature of one degree Celsius will lead to an increase of 18% in the number of cases of cutaneous leishmaniasis. Therefore, the model could be used by public health makers of the province in forecasting provided the availability of appropriate climate information and thus to be prepared to face the epidemic.

#### REFERENCES

- [1] F. Bachi, Aspects épidémiologiques et cliniques des leishmanioses en Algérie. La Lettre de l'Infectiologue - Tome XXI - N1 Janvier-Février 2006.
- [2] M. Chikhi, M. Chavance, Estimation du modèle linéaire généralisé et application, Sciences & Technologie A – N°36, Décembre 2012, 9-16.
- [3] A. Colin Cameron and Pravin K. Trivedi (2013), *Regression Analysis of Count Data*, 2nd edition, Econometric Society Monograph No.53, Cambridge University Press, 1998.
- [4] J. M. Hilbe, Negative Binomial Regression, second edition Cambridge University Press, United Kingdom 2011.
- [5] P. McCullagh and J. A. Nelder, Generalized Linear Models, London Chapman and Hall, 1989.
- [6] S. Selmane, Stability Analysis of a Human-Phlebotomus Papatasi-Rodent Epidemic Model, Interdisciplinary Topics in Applied Mathematics, Modeling and Computational Science. Springer Proceedings in Mathematics & Statistics, Vol. 117
- [7] Control of the leishmaniasis (2010) report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis, Geneva, , 22-26 March. (WHO technical report series ; no. 949).
- [8] Office national de météorologie. URL : <http://www.onm.meteo.dz>
- [9] Office national des statistiques. URL : <http://www.ons.dz>
- [10] Relevé épidémiologique mensuel, Institut national de la santé publique (INSP). <http://www.and.s.dz/insp/rem.html>.