International Journal of
Science and Engineering Investigations
vol. 6, issue 60, January 2017
ISSN: 2251-8843

Use of Generalized Estimating Equations in the Modeling of Cutaneous Leishmaniasis (Case: Algeria)

H. Elhadj[1], Y. Kerboua Ziari[2]
[1]High National School of Statistics and Applied Economics, Algiers, Algeria
[2]Faculty of Physics, University of Sciences and Technology Houari Boumediene, Algiers, Algeria
([1]elhadj.hayat@yahoo.fr, [2]yasminaziari@yahoo.fr)

Abstract-The application of statistical modeling in health field is an important tool as it is used for forecasting and measures the relationship between different diseases and factors that cause their emergence. Through this work we present a modeling of Cutaneous Leishmaniasis epidemic by taking climatic factors into account. We use for that Generalized Estimating Equations (GEE) and show the contribution of these models compared to other similar modeling approaches such as Generalized Linear Models (GLM) and Generalized Linear Mixed Models (GLMM).

Keywords- *statistical modeling, Cutaneous Leishmaniasis, GEE, GLM, GLMM*

I. Introduction

Public health has always been a sensitive sector given the impact it can have on economic and social development of the country, the good functioning of this sector concerns the quality of services but also the fight against dangerous diseases like some obligatory declaration diseases which have significant incidence rates in some regions in Algeria, as well as a spread towards new regions.

Zoonoses come first in obligatory declaration diseases, followed by waterborne diseases, then the meningitis and finally the other diseases. Cutaneous Leishmaniasis takes the most important part in terms of the incidences recorded by zoonoses, this vector disease is caused by the bite of an insect (Phlebotomus) having for its part been in contact with gerbils. Although this epidemic is generally linked to insecurity and precariousness in agricultural and water management programs in the affected areas, studies have shown that climatic factors have a direct impact on its incidence since Phlebotomus lives in warm areas, which reflects the presence of this epidemic in the steppe, arid and semi-arid regions, mainly in the northern fringe of the Sahara [H Elhadj 2015].

In order to measure the impact of climatic factors (mainly temperature) on the increase in the number of cases recorded by Cutaneous Leishmaniasis in the most affected regions we have used Generalized Estimating Equations.

II. Data and methods.

A. Data

The region concerned by the study comprises three regions of Algeria which have recorded a very high number of cases in recent years, there are provinces of Biskra, Msila and Eloued the number of cases recorded by the cutaneous leishmaniasis 'LC' were collected from the monthly epidemiological surveys published by the National Institute of Public Health (January2010-December2011). the climatic data for the same period grouping the temperature 'T' the humidity 'H', and the cumulative rainfall 'P' were obtained from the web site tutiempo.net [REM 2010/2011], [Tutiempo. net]

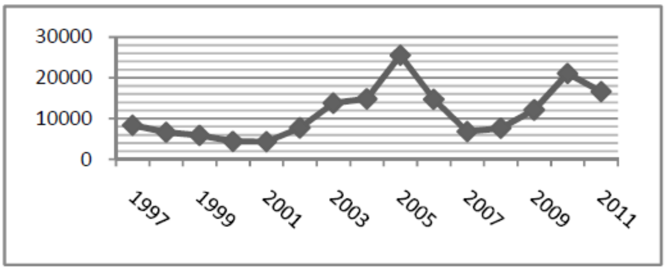

Figure 1. Evolution of the number of cases of cutaneous Leishmaniasis in Algeria during the Period 1997-2011

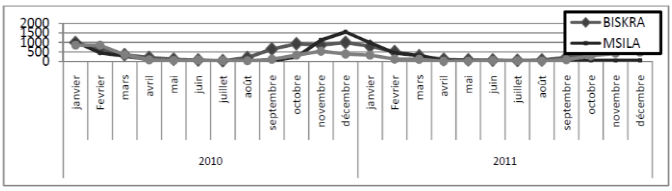

Figure 2. Evolution of the number of cases of Cutaneous Leishmaniasis in the wilayas of Biskra, Msila and Eloued in the period January 2010 - December 2011

The number of cases is very reduced during the hot months, this amounts to the fact that the adults appear towards the month of May, they are active during all the hot months, then disappear in the autumn. Furthermore, The incubation of this disease in humans is silent and lasts from two to six months, which explains the number of cases high in winter.

In Algeria, the oldest focus of the LC is Biskra (1980) before its extension towards 1982 M'sila in 1982 and Tiaret in 1985. The new affected regions are Oued, Ghardaia, Bechar et Laghouat and those of the North: Batna, Medea, Tiaret, Bordj, Bou Arreridj. It should be noted that the change in temperature and precipitation following climate change increase the risk of the spread of Cutaneous Leishmaniasis to other regions of Algeria. [Scientific exchange days 2012], [M'liki Feriel 2010].

B. Statistical Method

Classical linear regression models have recorded several extensions in order to be more adapted to the types of data.

When data are repeated over a number of individuals over time we talk about panel data, Extensions to this type of data include models Generalized Linear Models (GLM), which are used mainly to explain variables that do not follow the normal distribution.

The notion of fixed effects and random effects was introduced to provide more improvement in the modeling of panel data. The fixed effects model states that there is a dependence between individuals and time these effects are taken as unobserved variables correlated with the explanatory variables. This model also assumes an independence between the Y_{ij}, the indices i and j represent respectively the individual and the time [B S Ammour 2010].

The modeling of the correlation between the Yi which represent the values taken by the same individual in several consecutive periods of time which can intervene in the explanation of the phenomenon studied is sometimes necessary. Thus, random effects based on the assumption of the zero correlation between individual heterogeneity and the observed variables included in the model have been introduced into the generalized linear model to give rise to the Generalized Linear Mixed Models (GLMM) allowing the modeling of this correlation using the Maximum Likelihood (conditional approach), It assumes that the random effects follow a normal distribution. Moreover, and conditionally to the random effects, the endogenous variable follows an exponential distribution. The GLMMs therefore measure the effect of exogenous variables on the averages of each group.

As for the Generalized Estimation Equations (GEE), they are based on a marginal approach, they measure the effect on the marginal average of the population, in other words It does not fully specify the joint distribution of Y_i but give a modeling of the mean and a specification of the structure correlation.

The parameters of the model are convergent even if working correlation structure would be not well specified. In addition (GEE) use robust estimators for estimating model parameters and their variance-covariance matrix.

We assume that the density function of Y_{ij} given X_{ij} is:

$$f(y_{ij}/x_{ij}) = exp\left\{\frac{y_{ij}\theta_{ij}-b(\theta_{ij})}{a(\emptyset)} + c(y_{ij},\emptyset)\right\}, \quad i = 1,\dots\dots,n; \quad j = 1,\dots..n_i \tag{1}$$

Where $\mathrm{E}(y_{ij}|x_{ij}) = \mu_{ij} = g^{-1}(\eta_{ij}) = g^{-1}(x'_{ij}\beta)$, θ_{ij} Is a canonical parameter and represents the position parameter, $\emptyset$ is dispersion parameter, $x'_{ij}\beta$ Is the linear predictor, g(.)Is the link function.

The likelihood function is given by:

$$L(\beta,\emptyset,y_{ij}) = exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{y_{ij}\theta_{ij}-b(\theta_{ij})}{a(\emptyset)} + \sum_{i=1}^{n}\sum_{j=1}^{ni}c(y_{ij},\emptyset)\right\} \tag{2}$$

Consequently, the log-likelihood function is given by:

$$l(\beta,\emptyset,y_{ij}) = exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\frac{y_{ij}\theta_{ij}-b(\theta_{ij})}{a(\emptyset)} + \sum_{i=1}^{n}\sum_{j=1}^{ni}c(y_{ij},\emptyset)\right\} \tag{3}$$

In our case, and as we have count data poisson distribution is the most appropriate in modeling.

The first step consists in estimating the vector of the regression coefficients β which maximizes the likelihood function that the random variables of the vector Yi are independent (well that this hypothesis is generally false) by solving the system of equations:

$$\left\{\frac{\partial l(\beta,\emptyset,y_{ij})}{\partial\beta_k}\right\}_{k=1,\dots\dots,p} = [0]_{p\times 1} \tag{4}$$

This can be written in matrix form:

$$U_{indep}(\beta) = \sum_{i=1}^{n}\frac{1}{a(\emptyset)}X'_i\Delta_i(Y_i-\mu_i(\beta)) = 0 \tag{5}$$

Where $X_i = (x_{i1},\dots\dots x_{in_i})$', $x_{ij} = (x_{ij1},\dots.,x_{ijp})'$, $\mu_i(\beta) = \mu_{i1}(\beta),\dots\mu_{in_i}(\beta))'$, Δ_i Is a Diagonal matrix with dimension $n_i \times n_i$ whose element in position (j, j) $\frac{\exp(n_{ij})}{\{1+exp(n_{ij})\}^2} = \frac{\partial\mu_{ij}}{\partial\eta_{ij}}$.

The variance matrix of Y_i is:

$$V_i = Var(Y_i) = \emptyset A_i^{\frac{1}{2}}R_i A_i^{\frac{1}{2}} \tag{6}$$

GEE consists in making a generalization of equation (5) where we can assume a structure of correlation other than independence.

By writing $V_i = \emptyset A_i^{\frac{1}{2}}R_i(\alpha)A_i^{\frac{1}{2}}$ such that α represents the vector of the unknown parameters we try to find the true correlation structure of Y_i If we specify it correctly, we will have lower variance of estimators even in the case of a bad specification the inferences will be valid.

Estimation equations are given by:

$$\sum_{i=1}^{n}D'_iV_i^{-1}\{Y_i-\mu_i(\beta)\} = 0 \tag{7}$$

With $D_i = A_i\Delta_iX_i$

The estimator $\hat{\beta}$ which solves equation (7) is found by applying an algorithm that stops at convergence between $\hat{\beta}_{m+1}$ et $\hat{\beta}_m$ [BABACAR SECK, 2006]

Among the structures used in the correlation matrix

- Independent: in this case $R_i(\alpha)$ =$I_{n_i \times n_i}$. We do not need to estimate the parameter α

- Exchangeable $R_i = \begin{pmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \vdots & \vdots \\ \alpha & \alpha & \dots & 1 \end{pmatrix}$

- AR(1) $R_i = \begin{pmatrix} 1 & \alpha & \dots & \alpha^{n_i-1} \\ \alpha & 1 & \dots & \alpha^{n_i-2} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha^{n_i-1} & \alpha^{n_i-2} & \dots & 1 \end{pmatrix}$

III. RESULTS

Estimation of the GEE model was done using SPSS software were retained only the factor of temperature after removing other factors (Humidity and Rainfall) because of the non significance in the model. Table 1 shows the results of the estimation with the three forms of structure of the correlation matrix.

TABLE I. RESULTS OF GEE ESTIMATION

Structure of correlation	QIC	QICC	Parameter B	standard Error
AR(1)	13775,70	12872,47	0,07	0,0041
Exchangeable	13461,58	12207,59	0,093	0,126
Independent	13474,58	12197,79	0,094	0,128

It should be noted that the temperature was taken as a delayed variable of 6periods (months) since the incubation of the disease can last up to six months.

According to the results, the exchangeable correlation structure is the most suitable for the data (having the minimum values of the QICC criterion), in addition parameter B is estimated at 0.094, which means that the effect change of one unit of temperature implies a change of 9% in the average of number of CL cases recorded in all three regions of the study.

IV. CONCLUSION

Temperature remains the main climate factor influencing the incidence of Cutaneous Leishmaniasis, through this work we have attempted to measure this effect using the Generalized Estimation Equations.

Nevertheless, the integration of other non-climatic factors would be relevant to better explain the Presence (human factor, environment, etc.).

The fight against this epidemic in Algeria is carried out through the national anti-phlebotomy program. In view of the incidence of cutaneous Leishmaniasis and the risk of its spread to other regions it remains necessary to take more measures for a more effective fight against this disease.

REFERENCES

[1] BABACAR SECK : «Estimation pour les mod`eles linéaires généralisés: Approche marginale, approche conditionnelle et application», Faculté des Science et de Génie Université LAVAL de Quebec, Septembre 2006.

[2] Boubekeur Seddik Ammour: «Introduction des effets fixes dans un modèle binomial négatif», Mémoire présenté comme exigence partielle de la maitrise en Economie Université du Québec à Montréal 2010.

[3] Daowen Zhang: «Analysis of Longitudinal Data from Epidemiologic Studies», Graduate Summer Session in Epidemiology, NC State University.

[4] H. Elhadj, Y. KerbouaZiari, and S. Selman: Cutaneous Leishmaniasis Modeling: the case of Msila Province in Algeria, International Journal of Innovation and Applied Studies, Vol. 10 No. 1 Jan. 2015.

[5] Journées d'échanges scientifiques de l'Ouarsenis ,Prise en charge de la Leishmaniose en Algérie, Tissemsilt, Algérie, 26 mai 2012.

[6] M'liki Feriel : Impact des changements climatiques sur la Leishmaniose en Algérie, Mémoire de Magister, Université Badji Mokhtar-Annaba-, Faculté des science, Département de Biologie, 2010.

[7] Relevé Epidemiologique Mensuel, Institut Nationale de la Santé Publique.

[8] www.tutiempo.net