Use of a mixture statistical model in studying malaria vectors density

Olayidé Boussari[1], Nicolas Moiroux, Jean Iwaz, Armel Djènontin, Sahabi Bio-Bangana, Vincent Corbel, Noël Fonton, René Ecochard

1 : Université d'Abomey Calavi (UAC), LERSM/CIPMA/UAC 072 BP 50 Cotonou - Bénin

Keywords

Malaria, *Anopheles*, counting, heterogeneity, overdispersion, latent class mixture models, disease vectors

Introduction

Malaria is still a major public health issue in Sub-Saharan Africa. In 2010, this region bore 91% of the global disease death burden estimated to 655,000 deaths (World Health Organization). Studying the risk of vector transmission is at the basis of every survey about the importance of malaria in a given zone.

In 28 villages of Southern Benin, a recent cluster randomized controlled trial (RCT) aimed to compare the efficacy of different vector control interventions. In the study area, high variations in the density of malaria vectors were observed in time and space and there were many localities with zero mosquitoes collected during several nights.

In the present work, we assessed the ability of Poisson, Negative Binomial (NB), Zero Inflated Poisson (ZIP), Zero Inflated Negative Binomial (ZINB) and a Non-Parametric Mixture of Poisson (NPMP) models to fit the distribution of counts of malaria vectors measured in the 28 villages. Using a multivariate NPMP, we introduced a classification of the villages based on the mean vector density after adjustment for a set of environmental, demographic and climatic covariates. Then, we assessed the relationship between this classification and the vector interventions implemented in the villages. The results of this work will help design site-specific malaria vectors sampling.

Methods

Data collection: The data analyzed in the present study stem from mosquito collections carried out every 6 weeks between January and December 2009 (i.e. 8 surveys) in 28 villages of the sanitary region of Ouidah-Kpomassè-Tori (OKT) in South Benin. Only *Anopheles gambiae* and *Anopheles funestus* mosquito counts were considered in our work. The villages were divided into four groups (seven villages per group) where four different vector control measures were implemented (see Corbel et al.). Environmental, demographic and climatic data were also collected during the surveys.

Checking overdispersion and excess of zero in the data: The mean-variance relationship regarding the number of collected malaria vectors was analyzed graphically to explore data dispersion. We also assess the "excess of zero" through a graphical representation of the distribution of vector counts.

Approximation of the distribution of the data: The approximations of the distribution of the number of collected malaria vectors by the Poisson, ZIP, NB, ZINB and NPMP distributions were compared using the maximum likelihood (ML) estimation. A graphical representation of comparison results is used to show the counts as well as the predictions given by each of the above-cited distributions.

Multivariate Analysis: Given the hierarchical structure of the data collection system, another NPMP model was considered to allow for various components of the variance of the counts. In the latter model was adjusted on the environmental, demographic and climatic covariates. This model also allows for random intercept at "village" level and hence is a conditional model.

Assessing the impact of vector control strategies: For each village, a posteriori probability of belonging to each class after adjustment on all the covariates is estimated by the NPMP conditional model and each village is assigned to one of the classes based on the maximum of the a posteriori probabilities (MAP). Hence, the village grouping for implementation of the vector control strategies

and the classification resulting from the NPMP conditional model were compared using a Kruskall-Wallis test.

Results

Entomological Data: Total of 2,994 malaria vectors belonging to the *An. gambiae* complex and the *funestus* Group were collected during 3,584 human-nights of mosquito collection. The density of anopheline collected changed over space and time (Figure 1).

Study of the dispersion: Village, survey, and village-survey mean numbers of malaria vector collected per night on humans were plotted with their corresponding variances in Figures 2A, 2B and 2C respectively. The variances were much higher than the means; this indicates overdispersion of the data.

Distribution analysis: The cases for which zero malaria vectors were collected represented 74.7% of the total (Figure 2D). Figure 3 shows the bar diagrams and the expected density probability curves of the counts with the Poisson, ZIP, NPMP and NB models. The proportion of zero predicted by the Poisson distribution is significantly lower than that observed in the dataset. The curves relative to the NB and to the NPMP models are very similar and fit well the observed data distribution. Since the NB model allows for the excess of zeros, the proportion of zeros estimated by the ZINB model is nearly null; the NB and the ZINB models are therefore equivalent.

Multivariate Poisson mixture analysis: The NPMP conditional model with four latent classes shows that presence of market gardening, population density, mean cumulated rainfall over the 8 surveys, mean cumulated number of rainy days over the 8 surveys and outdoor position were positively associated with the number of malaria vectors caught on human. On the other hand, distance to a freshwater body, presence of water conveyance, presence of cattle, single-cluster village houses, mean NDVI and deviation at each survey from mean cumulated number of rainy days over the 8 surveys were negatively associated with the number of malaria vectors caught on human.

The 28 Villages were then classified according to their respective MAP. The mean number of malaria vectors collected ranged from 0.050 vectors per human per night in the 1st class (with only one village) to 0.713 in the 4th class (with 8 villages).

A Kruskal-Wallis test did not show a significant association between villages classification obtained from the model and the villages grouping for vector control strategies (Chi2 = 2.029, p-value = 0.566).

Discussion

McCullagh and Nelder asserted that whenever the variable of interest is a count, its distribution is often an overdispersed Poisson distribution. The present data are another illustration of this assertion.

In southern Benin, both spatial and temporal heterogeneities in vector densities were mentioned by Djènontin et al. (2010). With a NPMP conditional model, we found that this can be explained by some factors associated with the density of malaria vectors and our results agree with those found in most other studies.

In conclusion, we found that the NPMP model was useful to assess the relationships between vectors density and villages, climatic or environmental characteristics. It might therefore be an efficient tool to compute risk maps of the host-vector contact. Moreover, the NPMP model provided a classification of the villages after taking into account some covariates. Such a classification could be used at a pre-study step to improve the study design of mosquito collection and adapt the sampling effort according to the village characteristics, especially in region with high spatial and temporal heterogeneities of mosquito density, like in the OKT region. Furthermore, NPMP model could help in the study design of RCT when a stratified sampling is needed. The same model may be adapted and used in other settings for the study of the distribution of vectors of other diseases.

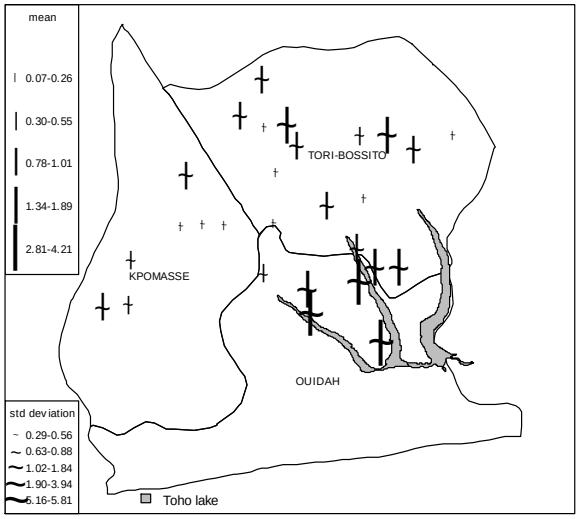

Figure 1 – Means and standard deviations of the number of mosquitoes collected per site and per night at each of the 28 villages of the study.

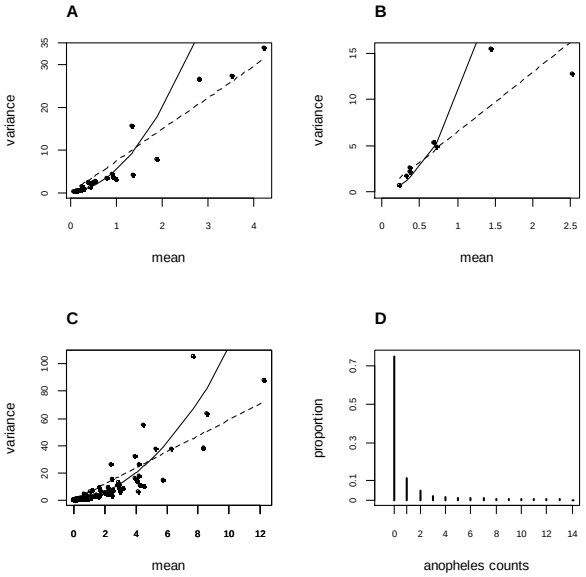

Figure 2 – Mean-variance diagrams of the number of malaria vectors collected per village (Panel A), per survey (Panel B), and per village-mission (Panel C). Panel D shows a bar diagram of the distribution of mosquito counts at each collection site (the scale of the X-axis was limited to 14).
On panels A, B and C: the dotted lines represent a linear link between the means and the variance ($\text{var}\,iance = \alpha \times mean$ with α = 7.4, 6.48 and 5.9 respectively); the curves represent a quadratic link between mean and variance ($\text{var}\,iance = mean + \beta \times mean^2$ with β = 4.4, 8.9 and 1.1 respectively).

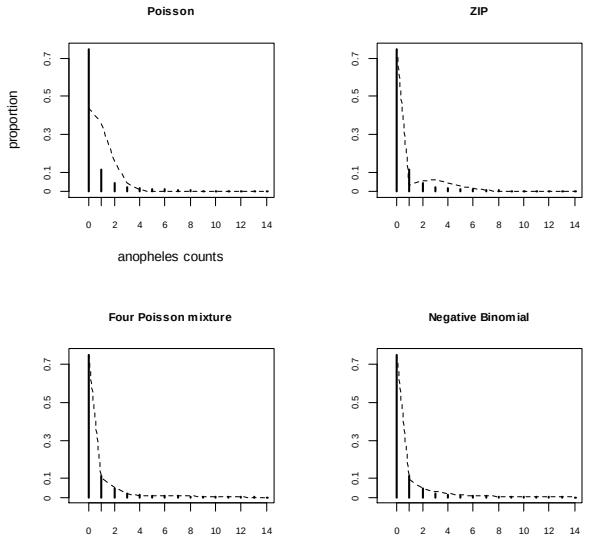

Figure 3 – Observed and expected proportions of mosquito counts according to Poisson, ZIP, NPMP and NB distributions.

PS: For more details see Boussari et al. (2012)