# Risk Mapping of East Coast fever in coastal and highland regions of Kenya based on predicted mortality and morbidity incidences.

C.M. Diaz,[1] S.C. Massawe,[1] A. Clemence,[2] G.K. Gitau,[3] H.K. Kiara,[1,] G.R. Muraguri,[4] C.J. O'Callaghan,[5] B.D. Perry [1] .

[1]International Livestock Research Institute, Nairobi, Kenya
[2,] Royal Veterinary College, University of London, UK
[3]African Union, Inter African Bureau of Animal Resources, Nairobi, Kenya
[4]Kenya Agricultural Research Institute, National Veterinary Research Centre, Muguga, Kenya
[5]Department of Community Health and Epidemiology, Queen's University, Kingston, Ontario, Canada

## SUMMARY

This paper proposes risk mapping as a tool for interpolating epidemiological data from intensively surveyed local sites to make conclusions over a larger area. Two ECF risk models were built using morbidity and mortality incidence data from the coastal and the central highlands of Kenya as dependent variables. It is concluded that the available data were not adequate to predict accurately mortality and morbidity of ECF. The main problem with the approach was its inability to capture the biophysical variability when the original data were distributed into grid-cells for spatial analysis.

## INTRODUCTION

East Coast fever (ECF) is a lymphoproliferative disease of cattle, caused by the protozoan parasite *Theileria parva*. The disease causes huge losses to farmers in Kenya, especially in the central highlands and the coastal region. Availability of adequate epidemiological data on the disease would aid prioritisation of interventions. However, these data are scarce or highly site specific and identification of recommendation domains for any intervention using conventional survey methods at national or regional levels is very expensive**.**

The objective of this study was to evaluate the use of risk mapping techniques as a tool for assessing ECF risk.

## MATERIAL AND METHODS

The analysis utilised data from 4 field studies and biophysical information. Field studies provided information on tick presence and abundance, tick control frequency and the main feeding system. These were combined with biophysical information such as agro ecological zonation, livestock census, land use and climatic data. Analysis involved the use of spatial and statistical tools. Epidemiological data were transformed in a 0.027 decimal degree grid-cell format (approximately 5 x 5 km), using the *spatial analyst* tool of Arc View software. Thus, the 5026 mortality and morbidity records were summarised in 72 grid-cells. Two different models were separately tested for predicting ECF mortality and morbidity having indices of disease risk and total number of cases analysed as the dependent variables. Logistic regression model was used for identifying significance among the surveyed and GIS derived data.

Calculation of the $R^2$ statistic, indicating correlation between the two sets of predictions, was used as a validation procedure of predictions based only on GIS derived information. The map calculator tool of Arc View software was used for assessing the level of ECF mortality and morbidity separately. The tool works by assigning a coloured pattern (cluster) to each of the 5 x 5 km grid-cells defining the Kenya territory.

**RESULTS**

From survey data, tick presence, regardless of species, tick control frequency and the main feeding system, were significant explanatory variables for ECF morbidity (dependent variable). Five dummies of AEZ, and PPE (a ratio from rainfall/evapotranspiration) for July were all significant explanatory variables from the GIS derived variables. Odds ratios and Wald Chi-square for all the explanatory variables are shown in Table 1.

**Table 1:** Analysis of Maximum Likelihood Estimates for ECF morbidity model

| Variable | DF | Parameter Estimate | Standard Error | Wald Chi-Square | Pr> Chi-Square | Standardized Estimate | Odds Ratio |
|---|---|---|---|---|---|---|---|
| INTERCEPT | 1 | -0.9109 | 0.4689 | 3.774 | 0.0520 | . | . |
| PPE July | 1 | -0.8987 | 0.2363 | 14.466 | 0.0001 | -0.2877 | 0.407 |
| Animal Density | 1 | -0.0196 | 0.00258 | 57.461 | 0.0001 | -0.5258 | 0.981 |
| AEZ-LM2 | 1 | -2.5329 | 1.0159 | 6.216 | 0.0127 | -0.3390 | 0.079 |
| AEZ-UM2 | 1 | 2.1710 | 0.4350 | 24.902 | 0.0001 | 0.3620 | 8.767 |
| AEZ-UM3 | 1 | 1.8332 | 0.4651 | 15.537 | 0.0001 | 0.1674 | 6.254 |
| AEZ-LH2 | 1 | 2.7709 | 0.9204 | 9.062 | 0.0026 | 0.0776 | 15.974 |
| AEZ-UM4 | 1 | 0.7884 | 0.2937 | 7.207 | 0.0073 | 0.0831 | 2.200 |
| R.Appendiculatus Presence | 1 | -1.9261 | 0.3976 | 23.469 | 0.0001 | -0.3324 | 0.146 |
| Tick Control Frequency | 1 | -1.0598 | 0.2222 | 22.746 | 0.0001 | -0.2966 | 0.347 |
| Main Feeding System | 1 | 0.8688 | 0.1643 | 27.950 | 0.0001 | 0.3062 | 2.384 |
| General Tick Frequency | 1 | 1.5502 | 0.3965 | 15.286 | 0.0001 | 0.2408 | 4.712 |

A cut-off value of 8% was selected for assessing presence and absence of disease. Sensitivity and specificity indices were 69.9% and 63.6%, respectively for the model. These levels were considered adequate for making ECF morbidity predictions at policy level.

The predictive equation was then transformed leaving GIS derived variables as the only independent variables in the model. In order to assess the accuracy of the predictions, a scattergram analysis was performed to compare predicted probabilities derived from predictive equations with and without surveyed variables. Predictions differed substantially ($R^2$ 0.135). Morbidity predictions derived from equations without survey variables are considered unreliable. Therefore, the hypothesis is refuted for ECF morbidity model (Figure 1).

The same procedure was applied for the ECF mortality model. Sensitivity and specificity indices were 68.5% and 52.3%, respectively for the model. The exhibited specificity value was considered too low for making predictions at policy level. Thus, no further attempt was made for extrapolating predictions.

## DISCUSSION

From the low values of sensitivity and specificity obtained from the models, it is concluded that predictions are inaccurate for use in intervention purposes. Specificity for the ECF mortality predictive model was especially low, making invalid any attempt for interpolating data outside the geographical limits of the surveyed areas. Sufficient biophysical variability was not captured by the 72 grid-cells analysed, a number considered too low for spatial modelling in a vast and highly diverse area such as the coastal and highland regions of Kenya. It might help to produce separate ECF risk maps for indigenous and exotic cattle, because they show different disease tolerance. Data analysis performed in this study could not differentiate risk by breed, because information from cross sectional studies did not show mortality and morbidity by breed. It might also help to undertake separate analysis for the coastal and highland regions, as climatic and parasite differences might render their combination invalid.

A map is presented as part of the results, purposely to show the risk inherent in new mapping techniques. The risk is because it is not easy to follow the statistical procedure behind a graphical representation. The danger is especially high when maps are used at policy level in developing countries with tight disease control budgets. However, a solid biological and statistical understanding of the whole risk mapping procedure is a prerequisite for verifying the source of information used and the methodology the disease patterns were graphically represented in a map.

Although both hypotheses were refuted, the present work was useful in demonstrating the problems encountered in utilising pre-existing epidemiological data for interpolating data in risk mapping exercises. This study is continuing.

**Figure 1: ECF risk map based on predicted morbidity incidence**