Epidémiol. santé anim., 1997, 31-32

HAEMORRHAGIC SEPTICAEMIA AND BLACK QUARTER FORECASTING MODELS FOR INDIA

Sudhindra[1], Yathinder P.V.[2], Rajasekhar M.[3]

La pasteurellose et le charbon symptomatique sont les plus répandues des maladies enzootiques du bétail indien. Les données épidémiologiques concernant 1821 enregistrements de 1987 à 1994 dans les 20 districts de l'Etat de Karnataka (Inde du Sud) ont été analysées par régression logistique pour développer un modèle de simulation prédictive de ces maladies. Les variables prises en compte (n=47) concernent la zone agro-écologique (10), des données météorologiques (9), temporelles (12), démographiques (5), sanitaires (1) et des informations relatives aux mouvements des animaux, à la commercialisation, aux pratiques de soins et autres facteurs associés (10). Le modèle de simulation a pris en compte uniquement les facteurs significatifs et leur contribution relative à l'occurrence ou la non-occurrence des maladies, soient les zones agro-écologiques et les données temporelles ou spatiales. Les zones agroécologiques représentaient un ensemble d'informations synthétiques sur les paysages intégrant les réponses écologiques aux macro-climats s'exprimant au travers des sols, de la végétation, de la faune et des systèmes aquatiques. Parmi les nombreuses variables explicatives, les zones agro-écologiques ont contribué pour 27 à 30% tandis que les autres contributions variaient entre 3 et 15%. Le modèle a reconnu une occurrence vraie de 71,71% pour la pasteurellose et de 76,21% pour le charbon symptomatique. A l'inverse, la non-occurrence vraie était de 81,24% et 78,18% respectivement. La spécificité et la sensibilité ont été de 61,06% et 87,50% pour la pasteurellose et 71,39% et 82,15% pour le charbon symptomatique. Il est intéressant de noter que la proportion de faux-positifs a varié entre 12,5 et 17,8% alors que le taux de faux-négatifs a varié entre 28,6 et 38,94% pour l'ensemble des modèles. Ces modèles prédictifs peuvent être améliorés ultrieurement en affinant le système de relevé des maladies et la prise en compte en temps réel des facteurs spatio-temporels et épidémiologiques à un niveau micro dans le pays.

India is endowed with varied ecological interactive systems that sustain huge and diverse livestock population maintained in different management, migratory and sanitary conditions. Several endemic bacterial, viral and protozoan diseases, which cause considerable economic loss to the rapidly growing livestock sector, presents a complex national epidemiological scenario with respect to their morbidity, mortality, short and long term trends and temporal and spatial aspects. Amongst endemic bacterial diseases, Haemorrhagic septicaemia (HS) ranks first and accounts for more than 40-55 % of mortality due to all other infectious diseases of bovines in the country (Khera,1979, Dutta.et al. 1990). During the period 1987-1994, in all 8,665 outbreaks of HS involving 100,293 attacks with a mortality rate of 31.5 per cent were recorded. In comparison, Black quarter (BQ), the second most important disease, accounted for 5,145 outbreaks and involved 97,868 attacks with a mortality rate of 15.7 per cent (Anon, 1994). Their morbidity, mortality, temporal and spatial patterns varied considerably. (Ramarao and Rao 1990). Keeping this in view, the HS and BQ forecasting models were developed based on location-specific national main and sub agro-climatic zones ('agcl') (90), surface meteorological (28), livestock demographic (5) and disease related (17) eco-pathological factors. The 'agcl' zones are the unique national landscapes of distinct ecological responses to macro climates as expressed by soil, vegetation, fauna and aquatic systems.

SOURCE OF DATA

HS and BQ profiles reported by the Government of India were used for developing simulated forecasting models. The national surface meteorological data was resourced from India Meteorology Department, Pune and Bangalore. The details of 'agcl' were collected from the ICAR national bureau of soil survey and land use planning, Nagpur. The baseline livestock demographic data was obtained from the Directorate of Animal Husbandry and Veterinary Services, Karnataka. The 1,821 records of occurrence or non occurrence of HS and BQ outbreaks pertaining to 84 monthly observations during 1987-1994 from 20 districts of Karnataka state were analysed in this study. These records took cognisance of the disease event as a bivariate dependant variable, namely the occurrence or non-occurrence of an outbreak with their corresponding 47 independent variables, which then prevailed in specific 'agcl' (10) of Karnataka state. The district-wise 'agcl'; calendar month number; number of villages; forest and irrigated areas; livestock markets and fairs; bovine, crossbred, sheep and goat population; veterinary clinics and milk producer co-operative units; morning vapour pressure and dew point; morning and evening humidity and temperature; wind speed; soil temperature; rainfall and number of rainy days in a month were included. Of these, 'agcl', month and soil temperature were categorical variables.

---

[1] [2] Scientists

[3] Project Co-ordinator, ICAR Project on Animal Disease Monitoring, Surveillance and Forecasting, IAHVB Campus, Hebbal, Bangalore- 560 024, India

STATISTICAL MODELLING

The logistic regression model, widely accepted as the standard method for regression analysis of dichotomous data in the biological fields (Hosmer and Lemeshow,1989) and the probability formulae were employed to develop the simulated forecasting models. The SPSS 6.00 (SPSS Inc. U.S.A), was used as application software.

OCCURRENCE OF AN OUTBREAK

Prob (outbreak) = $1/1+e^{z}$..(i)

where, z is the linear combination, $z = b_0 + b_1 x_1 + b_2 x_2 + ... + b_p x_p$ .b_0 and $b_{1...p}$ are coefficients estimated from the data; $x_{1...p}$ are the independent variables and 'e' is the base of the natural logarithm, approximately 2.718.

NON-OCCURRENCE OF AN OUTBREAK

Prob (no outbreak) = 1- Prob (outbreak)...(ii)

In these general purpose formulae, it was noted that the relationship between the independent variables and the probability of occurrence or non-occurrence of an outbreak were non linear. The probability estimates were always between 0 and 1, regardless of the value of z. In logistic regression, the parameters of the model were estimated using the maximum-likelihood method and the coefficients that matched the "most likely" observed results were selected.

HS FORECASTING MODEL

The 1,821 monthly records were initially scrutinised to remove misclassified records whose Studentized residuals were greater than two, there by retaining 1,774 for final analysis. Amongst the 47 independent variables, the analysis of variance provided regression coefficients, Wald tests of significance and partial correlation for only 29 factors (as shown in parenthesis below) which exerted significant influence on the occurrence or non-occurrence of HS outbreaks. The Z value of HS model was calculated using the regression coefficients and selected independent variables where: $z = b_0 + b_1$ (agcl) + b_2 (cal. month no.) + b_3(bovine population) + b_4 (crossbred population.) + b_6 (forest area) + b_6(evening humidity.) + b_7(irrigated area) + b_8(number of milk producers units) + b_9 (number of veterinary clinics). b_0 and $b_1 ... b_9$ are the respective regression coefficients. Occurrence and non-occurrence of outbreaks were then estimated by using the above general probability formula (i and ii).

BQ FORECASTING MODEL

The 1,821 monthly records were initially scrutinised to remove misclassified records whose Studentized residuals were greater than two, there by retaining 1,776 for final analysis. Out of 47 independent variables analysed, only 19 factors (as shown in parenthesis below) which exerted significant influence on the occurrence or non-occurrence of BQ outbreaks. The Z value of BQ model was calculated using the regression coefficients and selected independent variables as: $z = b_0 + b_1$ (agcl) + b_2 (morning humidity) +b_3 (no. irrigation tanks) +b_4 (no. livestock markets) + b_5 (number of rainy days in a month) + b_6 (number of villages) + b_7 (rainfall in a month) + b_8 (soil temperature) b_0 and $b_1 ... b_8$ were the respective regression coefficients. Occurrence and non-occurrence of outbreaks were then estimated by using the above general probability formula (i and ii).

ASSESSMENT OF THE FORECASTING MODELS

The quantitative results and the accuracy of HS and BQ forecasting model are presented in Tables I and II.

Table I
Results of HS and BQ forecasting models

Haemorrhagic septicaemia

	OB present	OB absent -	
PR +	PP=370	FP=146	P+=516
PR -	FN=236	PN=1022	P- =1258
	OB+ = 606	OB- =1168	Total 1774

Black quarter

	OB present	OB absent	
PR +	PP=564	FP=176	P+740
PR -	FN=226	PN=810	P-=1036
	OB+=790	OB- =986	Total1776

Based on these results, the following critical features of the forecasting models as compared to confirmed events were compiled (Table II), (Salman,1992). The inferences drawn are self explanatory.

Where PR+ = predicted positive result; PR- = predicted negative result; P+ = total predicted positive; P- = total predicted negative; OB = outbreak; OB+ = total outbreaks present; OB- = total outbreaks absent; PP = predicted positive; PN = predicted negative; FP = false positive; FN = false negative.

The qualitative assessment of the forecasting models was derived by a histogram of predicted occurrence and non-occurrence of the outbreaks against probability values of 0; 0.25; 0.50; 0.75 and 1. In the histogram, maximum events of non-occurrence and occurrence of HS or BQ polarised to the extreme left or right of the probability value of 0.5. However, 60 occurrence events each for HS and BQ were misclassified in the probability range of 0 to 0.25 in contrast to 40 of BQ and none of HS non-occurrences misclassified under probability range above 0.75. These patterns indicated a reasonably good fit of the models.

Table II
Critical features of forecasting models vs confirmed events

features	formula	HS	BQ	features	formula	HS	BQ
Sensitivity	PP / OB+	61.06 %	71.39 %	**Misclassification**	(FP+ FN)/Total	21.53 %	22.64%
Specificity	PN / Ob-	87.50 %	82.15 %	**Predictive value +ve**	PP / P+	71.71 %	76.21 %
Proportion of FN	FN / Ob+	38.94 %	28.60 %	**Predictive value -ve**	PN / P-	81.24 %	78.18 %
Proportion of FP	FP / OB-	12.50 %	17.80 %	**True prevalence**	OB + / Total	34.16 %	44.48 %
Accuracy	(PP+PN)/Total	78.47 %	77.36 %	**Apparentprevalence**	P+ / Total	29.09 %	41.67 %

DISCUSSION

The forecasting of livestock disease is a demanding task involving intricate balancing of the multifactorial components that go with the triad of host, pathogen and the environment. The dynamic forecasting model presented here has taken cognisance of these critical epidemiological parameters in a holistic manner.

For example, in HS model, 29 out of 46 independent variables were found significant (R^2 values) with the largest contribution of 27 and 11 % from the 'agcl' (10) and calendar months(12) respectively. The contribution of other seven factors ranged from absolute values of 5 -13 %. In the BQ model, 19 out of 46 independent variables were found significant with the largest single contribution of 30 % from 'agcl' (10) and nine others contributed to an absolute of 3 to 15 %.

The models recognised the true occurrences at 71.71 % for HS and 76.21 % for BQ. In contrast, the true non-occurrences were 81.24 % and 78.18 % respectively. The sensitivity and specificity were 61.06 and 87.50 % for HS and the corresponding percentages were 71.39 and 82.15 for BQ. Interestingly, the proportion of false positives ranged from 12.5 to 17.8% and false negatives ranged. from 28.60 to 38.94% for both the models. The level of misclassifications ranged from 21.53 to 22.64 % for HS and BQ predictions at the probability value of 0.5. These performance values would substantially improve if the probability cut off is revised below 0.5. Further, field validation studies are required to fine tune the forecasting models to establish the contribution of each of the location specific interactive components.

REFERENCES

Anon,.(1994). Annual report, ICAR project on animal disease monitoring, surveillance and forecasting, 1994, Bangalore 560 024, India.

Dutta, J., Rathore, B.S., Mullick, S.G., Singh, R. and Sharma, G.C. (1990). Epidemiological studies on occurrence of Haemorrhagic septicaemia in India. Indian Vet. J. 67. 893-899.

Hosmer, W.D. and Lemeshow, S. (1989). Applied logistic regression. John Wiley & Sons, New York.

Khera, S.S.(1979). The incidence and distribution of epizootic diseases in India. Bull. Off. Int. Epiz., 91, 331-347.

Ramarao and Rao, B.U.(1990), Trends and differentials in the incidence of Haemorrhagic septicaemia in Karnataka during 1979-1985, Indian Vet. J. 67, 1990: 990-907.

Salman M,. 1992, Properties of screening and diagnostic tests, Training course material for Veterinary. epidemiology, June 8-18, 1992, Fort Collins, Colorado, USA.