APPLICATION OF PRINCIPAL COMPONENT ANALYSIS FOR CLASSIFYING EAST COAST FEVER REACTIONS IN CATTLE CHALLENGED WITH *THEILERIA PARVA*.

G. J. ROWLANDS, A.J. MUSOKE, V.A. NENE, R.P. BISHOP, S.M. NAGDA, AND S.P. MORZARIA,*

SUMMARY

A statistically derived disease reaction index, based on daily parasitological, clinical and haematological measurements observed in 440 5-9 month old Boran cattle following laboratory challenge with *Theileria parva* in clinical trials of a sub-unit vaccine against East Coast fever, is described. Principal component analysis was applied to 13 variables including first appearance of schizonts, first appearance of piroplasms and first occurrence of pyrexia, together with duration and severity of these symptoms, and white blood cell count. The first principal component, which accounted for over 80% of the total variation expressed by the 13 variables, provided the definition for the disease reaction index, defined on a scale of 0-10. The extension of the method to 133 cattle exposed to natural tick challenge, for which incubation periods are unknown and white blood cell count is impracticable to measure, is also described. A correlation of 0.98 was found between the laboratory and field reaction indices.

INTRODUCTION

Clinical scoring methods have in recent years been applied in human medicine to provide generic methods for assessing health status in clinical trials. However, until recently (Rowlands et al., 2000) there appears to have been no published application of such methods in veterinary medicine. The publication by Rowlands et al. (2000) refers to experimental work undertaken to evaluate a *Theileria parva* sporozoite surface antigen (p67) as a potential sub-unit vaccine against the parasite in cattle (Musoke et al., 1992). *Theileria parva*, a tick-borne parasite, causes a disease known as East Coast fever (ECF). The disease is of immense economic importance in eastern, central and southern Africa and inhibits the development of livestock production (Mukhebi et al., 1992). In comparing responses between immunised and non-immunised animals, disease reactions have been principally based on a subjective interpretation by clinicians of the severity of observations on pyrexia, parasitosis and haematology and other clinical symptoms. Responses to infection have then been categorised into those for which there was either no reaction (NR) or only a mild reaction (MR) and those for which the reaction was moderate (MODR) or severe (SR) (Anonymous, 1989). Animals that died or were euthanased (humanely killed) were among those categorised as severe reactors.

* International Livestock Research Institute, P.O. Box 30709, Nairobi, Kenya

The reliance on a clinician's judgement of what constitutes a mild, moderate or severe reaction has led to uncertainty in the reliability of the final outcome. A more objective method was considered necessary to reduce the levels of subjectivity in defining the severity of reactions of cattle to ECF in vaccine trials. The method chosen was that of principal components. This is a statistical method that finds linear combinations of a set of measurements (in this case describing the extent and severity of parasitological and clinical symptoms) that maximise the variation contained within them, thereby displaying most of the original variability in fewer dimensions. The first principal component, which is the new variable that accounts for the highest proportion of the variation expressed by the original variables, often provides a general average representation of the contribution of each of the individual measurements. As will be shown later, this is what tended to happen here. The method has already been applied to 13 variables derived from daily parasitological, clinical and haematological measurements observed in 309 5-8 month old Boran cattle following various laboratory challenge experiments with *Theileria parva* (Rowlands et al., 2000). The first principal component, which was based on approximately equal contributions of the 13 variables, provided the definition of the disease reaction index, defined on a scale of 0-10.

Following several years of experimentation with needle challenge under laboratory conditions it was decided to evaluate the sub-unit vaccine in the field to assess its efficacy under natural tick challenge. The ECF reaction index in its existing form (Rowlands et al., 2000) was not directly applicable for field use since the individual incubation periods when cattle are exposed to natural tick challenge are unknown. Thus, the time to first detection of schizonts, one of the 13 variables used for the index, cannot be derived. Furthermore, determination of white blood cell count, another variable, is often impracticable under field conditions. The purpose of this paper is to describe the statistical approach taken in extending the method of Rowlands et al. (2000) to deriving a disease reaction index suitable for both laboratory and field use.

EXPERIMENTAL METHODS

Cattle

Boran cattle (*Bos indicus*), 5-9 months old, free of antibodies to *Theileria parva* by ELISA (Katende et al., 1998), were used. A total of 440 cattle challenged in 20 laboratory experiments and 133 cattle challenged in three field trials were available for calculating the ECF reaction index. The numbers of records available had thus increased from those used by Rowlands et al. (2000) in the initial development of the index.

Infection of cattle

Cattle were infected in laboratory experiments with Muguga stocks of *T. parva* stabilates. Each stabilate was inoculated subcutaneously over one of the parotid lymph nodes (local drainage lymph node) as described by Rowlands et al. (2000). Various immunisation regimens were used in these experiments, which led to the p67 formulation and treatment regimen evaluated in the field. Three field trials were undertaken in Kenya. Five laboratory experiments that used the same p67 formulation and treatment regimen as applied in the field are used to demonstrate the application of the reaction index in assessing efficacy of immunisation.

Monitoring of cattle

Rectal temperatures of cattle in both field and laboratory experiments were recorded daily after challenge. On day 5 after inoculation of cattle in laboratory experiments, and at daily intervals thereafter, needle biopsy smears were made from the local drainage lymph node, stained with Giemsa's stain and examined for the presence of schizonts. The degree of parasitaemia detected was scored on a scale from 1 to 3. Biopsy smears were similarly taken daily from the contralateral, pre-scapular lymph node from the day after the draining lymph node was found to be positive with schizonts. Both parotid lymph nodes were examined in cattle exposed to field challenge from day 5 after exposure. As soon as schizonts were detected in one of the lymph nodes this was designated as the local drainage node. Thereafter, biopsy smears were taken from both lymph nodes

Blood smears, prepared from the ear vein, were collected daily from the day after an animal first became positive with schizonts. These smears were stained with Giemsa's stain and piroplasm parasitaemia was determined as the number of infected erythrocytes per 1000. For cattle receiving a laboratory challenge blood for determining total white blood cell count (WBC) was taken from the jugular vein in EDTA before infection and three times per week thereafter.

Clinically severe cases of ECF

Cattle which became severely affected with theileriosis, the predominant signs of which included long durations and high levels of pyrexia and parasitosis, low white blood cell counts and poor general body condition, appetite and respiration rate, were humanely killed. Animals less severely affected were allowed to recover or, in one of the field trials, treated with an anti-theilerial drug.

PRINCIPAL COMPONENT ANALYSIS

Laboratory experiments

A number of variables were derived from the parasitological, haematological and rectal temperature measurements taken (Table 1). One hundred and twenty six animals inoculated with a laboratory challenge were either killed on the basis of parasitological, clinical and haematological signs, or died, between day 14 and 22 post challenge. Thus, the lengths of the periods over which measurements were taken were shorter for these animals than for animals that recovered. In order to allow all animals to be compared on an equal basis, the period of measurement was fixed to 23 days, both for animals that died or were killed, and for animals that recovered. Thus, records were extended to day 23-post inoculation for any animal that did not survive that long. In doing so, the assumption was made that symptoms apparent at the time of death would have persisted at the same average intensity had animals survived until day 23. When an animal's temperature had returned to normal by the time of death, however, duration of pyrexia was not extended.

Table 1. Variables defined from measurements made on 440 5 to 9-month-old Boran cattle challenged in laboratory experiments and 133 challenged in field experiments for defining an ECF reaction index.

Abbreviation	Description	No. of records available	
		Laboratory experiments	Field experiments
SC1-FST	First day post inoculation that schizonts were detected in the local drainage lymph node.	358	0
SC1-LEN	Length of period (days) over which schizonts were observed.	232	113
SC1-INT	Average score intensity over the period that schizonts were observed.	358	129
SC2-FST [a]	First day post SC1-FST that schizonts were detected in contralateral, pre-scapular lymph node.	311	73
SC2-LEN	As for SC1-LEN	185	57
SC2-INT	As for SC1-INT	311	73
TEMP-FST [a]	First day relative to SC1-FST that a temperature > 39.4°C was recorded.	354	109
TEMP-LEN	Length of period (days) over which pyrexia was observed.	228	93
TEMP-INT	Average intensity of pyrexia, expressed as deviation from 39.4°, over the period that pyrexia was observed.	354	109
PIRO-FST [a]	First day post SC1-FST that piroplasms were detected.	324	88
PIRO-LEN	Length of period (days) over which piroplasms were observed.	198	74
PIRO-MAX	Maximum value recorded.	324	88
WBC	Mean white blood cell count between day 13 and day 19-post inoculation.	440	0

[a] The definitions of SC2-FST, TEMP-FST and PIRO-FST are given as deviations from SC1-FST rather than as days post inoculation, which are slightly different from those of Rowlands et al. (2000).

In cases where either schizonts, piroplasms or pyrexia were not observed (for example, 73 animals that demonstrated no parasitosis or pyrexia, and 47 animals that were detected with schizonts in the local drainage and not the contralateral, pre-scapular lymph node) values were substituted for missing days of first occurrence. Thus, a value of 14 for days post challenge was substituted for 'first detection' of schizonts in the local drainage lymph node (SCI-FST). Similarly, values of 8, 6 and 10 were substituted for the subsequent number of days to 'first detection' of schizonts in the contralateral, pre-scapular lymph node, for pyrexia and for piroplasms, respectively, following SC1-FST. Frequency distributions showing the variations in the first day of occurrence of schizonts, piroplasms and pyrexia are shown in Fig. 1. The above definitions were based on the observation that the majority of observed reactions to infection had taken place by the respective time point. In each case when symptoms were not recorded their duration was defined to be zero.

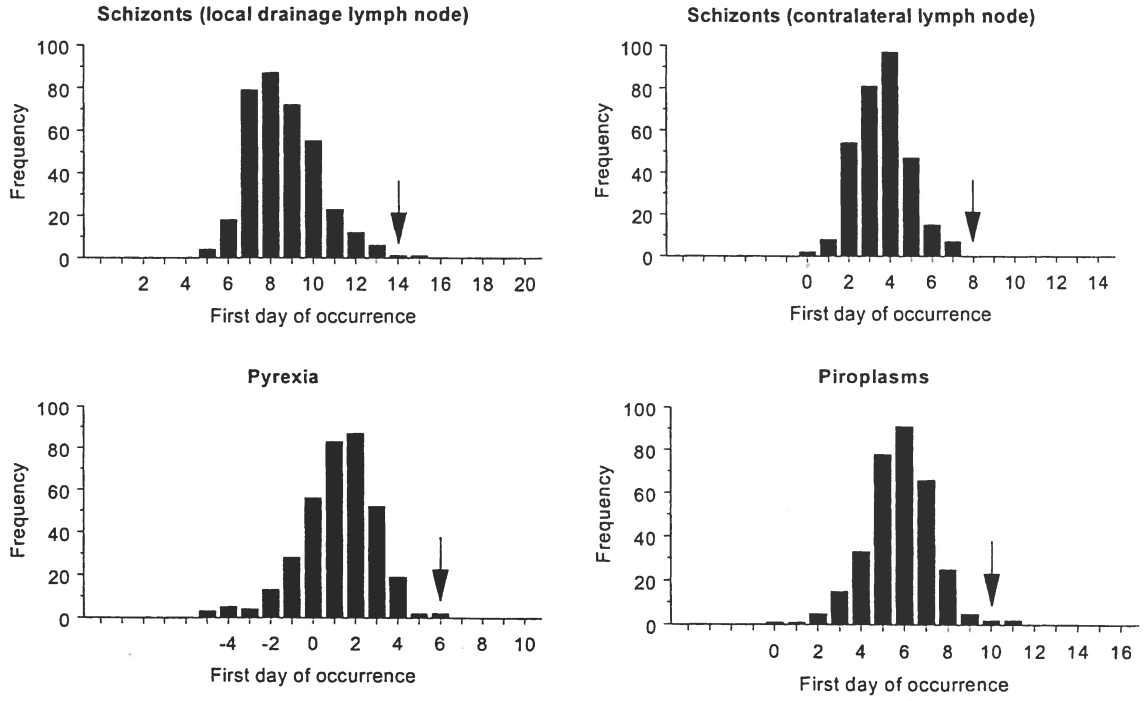

Fig. 1. Histograms of the first days of detection of schizonts in the local drainage node and the number of subsequent days to first detection of schizonts in the pre-scapular, contralateral lymph node, to first detection of pyrexia and to first detection of piroplasms in Boran cattle. The vertical arrows indicate the values substituted for animals that showed no symptoms.

Mean white blood cell count was calculated between day 13 and 19 when minimum counts were generally reached. Mean pyrexia was calculated as the mean deviation (rectal temperature – 39.4°C) over the period for which pyrexia was observed. A rectal temperature of 39.4°C or below during this period contributed a zero value in the calculation of the mean intensity of pyrexia. Similarly, failure to detect schizonts on a particular day during the period that schizonts were recorded counted as zero in the calculation of the average schizont intensity. The maximum piroplasm parasitaemia per 1,000 erythrocytes recorded was transformed to natural logarithms to make the data less skewed.

The 13 variables were then standardised $\left[x' = \left(x - \bar{x}\right)/sd\right]$ before subjecting the data to a principal component analysis using Genstat (Payne et al., 1993). For the purposes of this analysis only 232 animals that recovered and showed some clinical reaction were used, thus excluding observations based on statistically extended periods for animals that were humanely killed or animals that died, and animals that showed no clinical reactions (NR) at all. The method was then repeated on these 232 animals, but without SC1-FST and WBC, variables not possible to determine under field conditions

Field trials

Periods of recording of parasitological and clinical measurements obtained from the field were fixed to 15 days for calculation purposes. This period was based on the assumption that schizonts were first detected in the local drainage lymph node at an average of 8 days post challenge (see Fig. 1). This gives a total period of observation post challenge of 23 days, the same as that used for the analysis of the laboratory data. Other definitions were as described for the laboratory experiments. Ninety animals that recovered from infection under field challenge and showed some clinical reaction were available for verification of the index under field conditions.

RESULTS

Principal components

Table 2 shows the coefficients of the first principal component when applied to data assembled from both laboratory and field experiments. When all 13 variables were used the component represents an overall 'average' contribution of the six variables representing length and intensity of infection. There are smaller contributions from coefficients representing time to first detection of schizonts and subsequent time to detection of schizonts in the contralateral, pre-scapular node and of piroplasms. The time of onset of pyrexia relative to that of schizonts was of negligible importance. When restricted to the 11 variables that can be determined under field conditions similar coefficients were obtained. Similar values were also obtained when the method was applied to the parasitological and clinical observations determined from 90 cattle under field challenge. This verifies that the principal components derived from disease indicators for cattle under laboratory challenge are applicable to those under natural field challenge. Furthermore, when all animals, including non-reactors and animals that were humanely killed or died, were considered, Rowlands et al. (2000) found that a slightly higher weight was attached to SC1-FST, but that otherwise coefficients were similar. This confirmed that the assumptions made in substituting values for non-reacting animals and animals that were humanely killed or died were reasonable.

The coefficients obtained from the first two columns of Table 2, based on the analyses of 13 and 11 variables derived from the laboratory data, were then applied to all cattle in laboratory and field experiments, respectively. The resulting scores were then standardised to ensure that they lay in a range between 0 and 10. When similar indices for the 440 animals recorded under laboratory challenge were calculated using just 11 variables, their correlation with those derived using all 13 variables was 0.98. This indicates that little information was lost when dropping WBC and SC1_FST from the analysis.

Table 2. Coefficients of the first principal component in a principal component analysis of pyrexia, parasitological and haematological measurements in 5 to 9-month old Boran cattle.

	Laboratory experiments		Field experiments
	All 13 variables (232) [a]	Without SC1-FST, WBC (232)	without SC1-FST, WBC (90)
SC1-FST [b]	-0.18	-	-
SC1-INT	0.30	0.32	0.24
SC1-LEN	0.32	0.33	0.31
SC2-FST	-0.25	-0.28	-0.32
SC2-INT	0.31	0.32	0.34
SC2-LEN	0.34	0.35	0.36
TEMP-FST	-0.04	-0.06	-0.14
TEMP-INT	0.31	0.32	0.28
TEMP-LEN	0.28	0.29	0.27
PIRO-FST	-0.20	-0.23	-0.30
PIRO-MAX	0.34	0.35	0.34
PIRO-LEN	0.32	0.34	0.34
WBC	-0.26	-	-
Total variation accounted for (%)	55	61	53

[a] Numbers of records used in data analysis – these were for animals that demonstrated a clinical reaction to infection and recovered.

[b] For description of variables see Table 1.

Analysis of five laboratory trials

The individual reaction indices for animals under laboratory challenge undergoing the same immunisation regimen as applied in the field trials are illustrated in Fig. 2. These indices are shown as the first principal component along the x-axis, plotted against scores determined for the second principal component along the y-axis. As discussed by Rowlands et al. (2000), this second principal component provided no meaningful biological interpretation; it is useful though for the purposes of displaying the data in the form of a scatter diagram. Inspection of these scatter diagrams shows that those animals that died or were humanely killed tended to have high index scores, whereas those that recovered tended to have low index scores. There was some overlap, however, in the middle. Animals with high index scores that recovered are nevertheless animals that will have developed sufficiently severe clinical symptoms likely to warrant chemotherapy if they were raised on a farm. A score of 6, based on a general assessment of reactions as determined by the clinicians (see Fig. 2), was defined to separate animals showing

signs of severe ECF from those showing mild signs of the disease. An animal with severe ECF was thus considered to be an animal that would die from exposure to the disease or would develop a sufficiently severe reaction (≥ 6) likely to warrant chemotherapy.

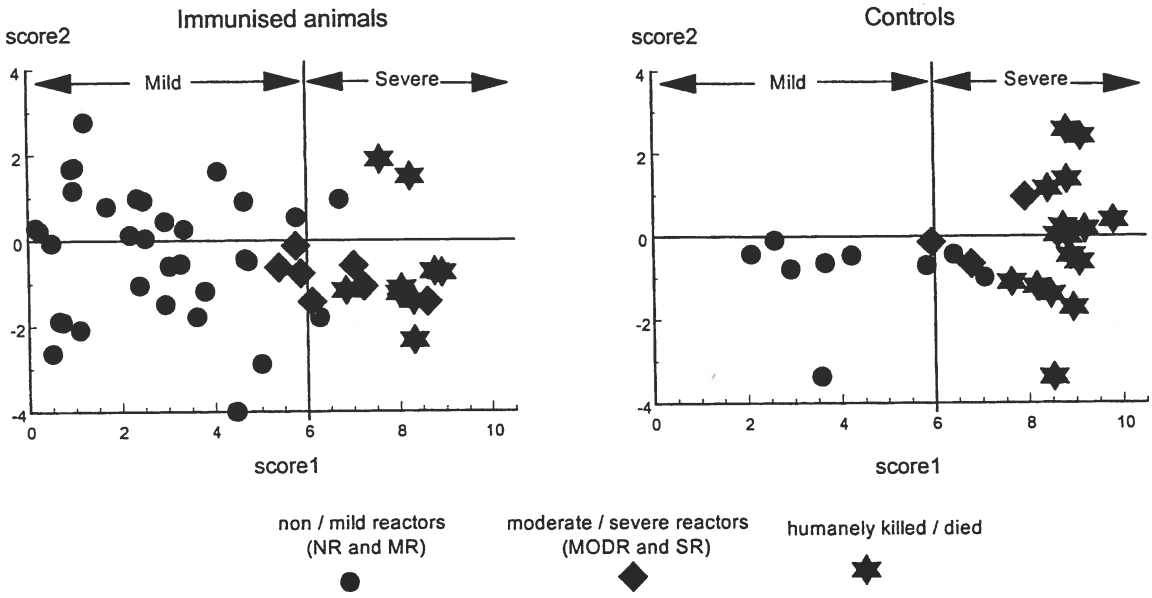

Fig. 2. Scatter diagrams of the first and second principal component scores for Boran cattle, susceptible controls or immunised with a p67 sporozoite antigen in five laboratory trials. The vertical lines separate statistically defined mild and severe cases of ECF. The symbols indicating severity of reaction refer to clinical assessment by the clinicians.

The number of mild and severe reactions defined in this way are summarised in Table 3 and the incidences of severe ECF in controls and immunised animals in each of the experiments are shown in Fig. 3. Application of logistic regression analysis to these data demonstrated reductions in proportional incidence from 0.80 ± 0.07 in susceptible controls to 0.39 ± 0.06 in immunised animals. The corresponding odds ratio, with 95% confidence limits, was 0.13 (0.04 – 0.43).

The data from the field trials have been analysed in the same way and await publication.

Table 3. Numbers of statistically defined mild and severe ECF reactions among controls and animals immunised with the sub-unit p67 vaccine in five laboratory experiments.

	Mild ECF	Severe ECF	Total
Controls	6	29 (25) [a]	35
Immunised	36	20 (14)	56

[a] Numbers in parenthesis are animals that died or were humanely killed.

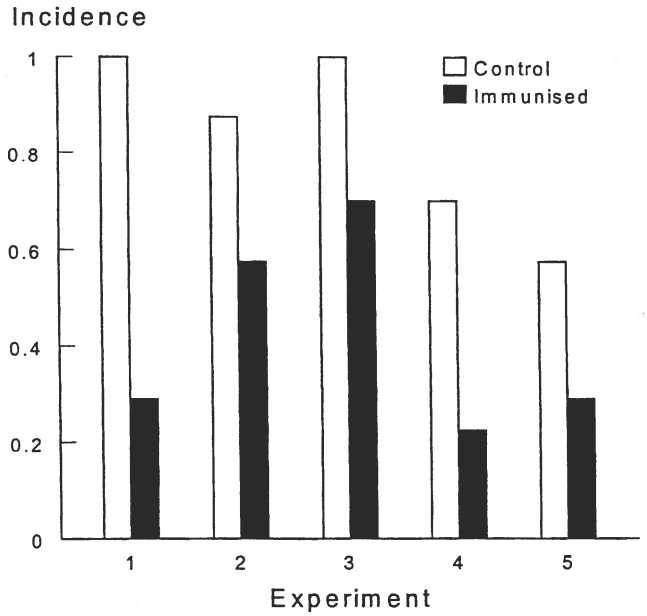

Fig. 3. The incidence of severe ECF in non-immunised and immunised Boran cattle in five laboratory experiments.

DISCUSSION

Principal component analysis is only one of a number of approaches that have been adopted in human medicine – others include factor analysis, discriminant analysis and other special methods. A brief review is given by Rowlands et al. (2000). The major advantage of any of these indices is that they reduce the element of subjectivity implicit in clinicians trying to formulate a balanced view of the severity of infection, in our case from the various parasitological, clinical and haematological symptoms observed. It is important to emphasise, however, that the index, as developed in our situation, facilitates statistical analysis of trial results once an experiment has been completed, but cannot be used as an aid to clinicians. In our example the data were categorised into cases of severe and mild ECF and analysed by logistic regression. This was done to determine levels of protection provided by the vaccine.

Having developed our index score, however, its use as a continuous variable in an analysis of variance is also possible. In small-scale laboratory experiments, where a qualitative approach to defining severity of disease reaction may yield non-significant results, statistical analysis based on the value of the score itself improves the power of the statistical comparison between immunised and non-immunised groups. This does not mean that sample size can be reduced, since mortality or disease severity will remain an important trait for assessing vaccine efficacy, but rather that the data collected can be more efficiently analysed.

It is reassuring that the statistical method assigned higher index scores to those animals that were considered necessary to kill than to some of the animals that died (Rowlands et al., 2000). This would indicate that in the majority of cases clinical judgements were sound, and that delaying further the decision to sacrifice an animal would have caused animals unnecessary suffering. In contrast, in the few cases where there was overlap with cattle that recovered euthanasia may have been premature. It is furthermore reassuring that good correlation was

found between the statistically derived and clinically defined disease reactions (Rowlands et al. 2000).

Although the first principal component accounted for only 50 to 60% of the total variation expressed by the original variables, this analysis was done on only the subset of those animals that recovered from infection. As shown by Rowlands et al. (2000) the proportions of variation accounted for increased to over 80% when all animals were considered, including those that were humanely killed or died. It is also worthy of note that the indices based on 11 and 13 variables were highly correlated. This suggests that even under laboratory conditions evaluation of WBC may be unnecessary.

In conclusion, the levels of protection of the p67 vaccine determined in the large number of laboratory trials (Musoke et al., 1992) can be concluded to have been based on sound and unbiased clinical judgement. Indeed, use of the statistically derived reaction index has not altered the conclusions that could have been made by clinical judgement alone on the efficacy of immunisation. Rather, the index has provided a degree of objectivity in the reporting of results and a level of credibility in their acceptance that will be crucial for the presentation of the field trial results.

REFERENCES

Anonymous (1989). Theileriosis in eastern, central and southern Africa. Proceedings of a Workshop on East Coast Fever Immunisation, Lilongwe, Malawi, 1988. Appendix 2, pp. 187-188.

Katende, J.M., Morzaria, S.P., Toye, P., Skilton, R.A., Nene, V., Bishop, R.P. and Musoke, A.J. (1998). The identification of a diagnostic antigen and the development of an ELISA for *Theileria parva* using a specific parasite immunodominant recombinant antigen. Parasitol. Res. 84, 408-416.

Mukhebi, A.W., Perry, B.D. and Kruska, R. (1992). Estimated economics of theileriosis control in Africa. Prev. Vet. Med. 12, 73-85.

Musoke, A., Morzaria, S., Nkonge, C., Jones, E. and Nene, V.A. (1992). Recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. Proceedings of the National Academy of Sciences, USA 89, 514-518.

Payne, R.W., Lane, P.W., Ainsley, A.E., Bicknell, K.E., Digby, P.G.N., Harding, S.A., Leech, P.K., Simpson, H.R., Todd, A.D., Verrier, P.J., White, R.P., Gower, J.C. and Tunnicliffe Wilson, G. (1993). Genstat 5 Release 3 Reference Manual. Clarendon Press Oxford.

Rowlands, G.J., Musoke, A.J., Morzaria, S.P., Nagda, S.M., Ballingall, K.T. and Mckeever, D.J. (2000). A statistically derived index for classifying East Coast fever reactions in cattle challenged with *Theileria parva* under experimental conditions. Parasitol. 120, 371-381.