Japanese Journal of Veterinary Research 65(3): 127–133, 2017

JJVR

REGULAR PAPER Regional Study

Discriminant analysis for the prediction and classification of tick-borne infections in some dairy cattle herds at Dakahlia Governorate, Egypt

Eman A. Abo El Fadl[1, §], Maged El-Ashker[2, §], Keisuke Suganuma[3, 4] and Mitsunori Kayano[3, *]

[1] Department of animal husbandry and development of animal wealth (biostatistics), Faculty of Veterinary Medicine, Mansoura University, Mansoura 35516, Egypt.
[2] Department of Internal Medicine and Infectious Diseases, Faculty of Veterinary Medicine, Mansoura University, Mansoura 35516, Egypt.
[3] Research Center for Global Agromedicine, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho, Obihiro, Hokkaido, 080–8555, Japan
[4] National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho, Obihiro, Hokkaido, 080–8555, Japan

Received for publication, December 26, 2016; accepted, April 25, 2017

Abstract

This study was undertaken to use the variable loadings in linear discriminant analysis (LDA) to determine the most important predictors for the discrimination of tick-borne diseases (TBDs), particularly babesiosis and anaplasmosis and predict the group membership from the predictors. In total, 163 cattle, from different localities at Dakahlia Governorate, Egypt, were investigated in 2012 and 2013 for the presence of TBDs. All cattle were clinically examined and a clinical index score was determined for each cow. Blood samples were also collected from each animal for adopting microscopy and diagnostic laboratory methods. Out of the examined cattle, 83 animals were acutely-ill (*Babesia bovis* and *Anaplasma marginale* were identified in 11 and 10 animals, respectively), while 80 cows were apparently healthy but having previous attacks of blood parasites (23 animals harbored *anaplasma marginale* (asymptomatic carriers)). The remained 119 animals were negative to TBDs. Fourteen animals were not survived and 149 cases were survived. As the result of the first LDA to discriminate babesiosis, anaplasmosis and negative to TBDs, 89.0% of animals were correctly classified; 78.8% (26/33) for *anaplasma*, 100% (11/11) for *babesia* infections, 90.8% (108/119) for negative to TBDs, respectively. The important predictors for the discrimination were oculonasal discharge, bloody feces, hemoglobinuria, bloody feces and respiratory rate. On the other hand, the second LDA discrimination showed high classification accuracy of 87.1% for the discrimination of survivors and non-survivors; 89.9% (134/149) for survivors and 57.1% (8/14) for non-survivors, while the important predictors included oculonasal discharge, recumbent posture and nervous sign.

Key Words: Classification; discriminant analysis; pathogens; predictors; tick-borne

§Equally contributed

*Corresponding author: Mitsunori Kayano, Research Center for Global Agromedicine, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho, Obihiro, Hokkaido, 080–8555, Japan
Phone: +81–155–49–5521. Fax: +81–155–49–5593. E-mail: kayano@obihiro.ac.jp
doi: 10.14943/jjvr.65.3.127

Introduction

Tick-borne diseases (TBDs), caused by *Babesia bovis, B. bigemina* and *Anaplasma marginale,* are crucial infections of cattle in tropical and sub-tropical regions in Egypt[1,2,4].

These infections cause constraints to the livestock improvement programs in Egypt with resulting a decrease of animal productivity and extreme economic losses. Animals suffering from acute attacks of babesiosis or anaplasmosis can pertain similar clinical presentation. Although the associating clinical symptoms are not pathognomonic, animals with chronic infections can be asymptomatic carriers[1,9,11]. Hence, accurate diagnosis as well as the prediction of the clinical outcome of these infections is required for appropriate and immediate treatment after the infection.

Discriminant analysis (DA) is a statistical technique used for differentiating groups (categorical dependent variable) by using the various independent and typically quantitative variables (parameters). Linear discriminant analysis (LDA) is mainly used in statistics (1) to find a linear combination of variables that classifies objects or events into two or more classes and (2) to identify important variables for the classification[8,14,15]. LDA allows us to predict a group membership based on a linear combination of the independent variables, hence careful examination of the prediction model that results from the procedure can give an insight into the relationship between group membership and the variables used to predict group membership[16].

In previous reports, it has shown that DA could be used for discriminating the type of pathogens causing infection and to determine the best predictors for classification. It had been used to analyze data of eleven variables involving (red blood cells, packed cell volume, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin content, hemoglobin concentration, white blood cells, neutrophils, leukocytes, platelets, rectal temperature), to identify which parameters out of the eleven are the most reliable parameters to be considered as markers of the ovine anaplasmosis[3].

On the other side, the forward stepwise discriminant function analysis (DFA) have been used for classification of the groups of Atlantic cod (*Gadusmorhua*) by using several parasites as "parasite tags"[6]. In another report, DA on the morphometric data of opisthaptor to determine which parameters more accurately distinguish between *Diplectanum cazauxi* and *Diplectanum bauchotae* has been reported and indicated that the best discriminant parameters are the copulatory apparatus and the central bar which determined the absence of misclassification in the model[7].

Up to now, there have been limited data about the value of using discriminant analysis in cattle with TBDs. In line with these considerations, the present study was delineated to check the significance of different determinants to discriminate the type of infection and to predict the clinical outcomes of the affected subjects.

Materials and Methods

Animal population and data collection: The present study included 163 cattle (130 from 6 dairy farms; 33 small-holders) that were located at Dakahlia Governorate, Egypt. The original study included 165 animals, however two animals were not follow the study design and thus were excluded. The animals were investigated for the presence of one or more TBDs during summers 2012 and 2013. The ages ranged between 1–5 years in dairy farms and between 6 months and 2-years for the animals of small-holders. The 83 animals, of four dairy farms (n = 50) as well as those of small-holders (n = 33), had recent attacks of babesiosis or anaplasmosis with variable degrees of tick infestation as well as sporadic cases of sudden deaths in the respective herds (acute cases). These cattle typically had pyrexia, pale to icteric visible mucosa, respiratory

embarrassment, and oculonasal discharge. Some rare cases showed enlarged superficial lymph nodes and the others had red urine. In the remained two farms (n = 80), no clinical signs appeared upon examination but a previous outbreak was the main complaint of the enclosed owners (chronic cases). Out of the 83/163 acute cases, 11 and 10 animals were infected with babesiosis and anaplasmosis respectively, while 23 chronic animals out of 80/163 were positive for anaplasmosis. Fourteen infected animals were not survived, and 149 cases were survived.

All cattle were clinically examined and a clinical index score was determined for each cow. The scoring system was based on clinical and laboratory variables including rectal temperature, heart rate, respiratory rate, coughing, dyspnea, mucous membrane color, superficially located lymph nodes, oculonasal discharge, color of urine, appetite, blood in feces, nervous signs, posture of the animal, frothy salivation and packed cell volume. Microscopic examination of Giemsa stained blood smears was performed for all investigated cattle.

Research hypothesis:

H_o: Any linear combination of the predictors has no discriminating ability to tick-borne disease.

H_A: There exists at least one linear combination of the predictors has discriminating ability to tick-borne disease.

Statistical analysis:

Discriminant analysis (DA)

DA is a statistical multivariate technique used in many different fields. LDA involves a discriminant variety and represents a linear combination of two or more predictors that discriminate between the objects in the groups defined a priori. The assumptions of LDA can be summarized as: Homogeneous within the group variances, multivariate normality within group, and linearity among all pairs of variables, no multi-collinearity was detected [10].

Data were analyzed by using SPSS version 16, USA. Two LDA models were used. The first model was carried out to check the significance of different determinants to discriminate the type of TBDs (*Anaplasma, Babesia*) as dependent variables by using clinical signs as independent variables. The type of infection (dependent variable) was coded before starting the analysis as following (*Anaplasma* = 1, *Babesia* = 2 and negative to TBDs = 3). All tested clinical findings including rectal temperature, heart rate, respiratory rate, cough, dyspnea, mucous membranes color, enlarged lymph nodes, oculonasal discharge, hemoglobinuria, appetite, blood feces, nervous signs, posture, frothy salivation were then entered in DFA model using stepwise method and from all of these variables only respiratory rate, hemoglobinuria, bloody feces, oculonasal discharge, mucous membrane and size of lymph node were selected to be in the equation.

The second LDA was carried out to check the significance of different determinants to predict the clinical outcomes of the affected subjects (survivors vs. non survivors) using clinical outcomes as dependent variables and the clinical findings as independent variables. The dependent variable the affected subjects was coded as following (survivors = 1 and non survivors = 0). All examined clinical signs including rectal temperature, heart rate, respiratory rate, cough, dyspnea, mucous membranes color, enlarged lymph nodes, oculonasal discharge, hemoglobinuria, appetite, blood feces, nervous signs, posture, frothy salivation were then entered in DFA model using stepwise method and from all of these variables only oculonasal discharge, posture and nervous signs were selected to be in the equation. As described in the first LDA model. For both models, Wilk' lambda, group centroids and classification accuracy were calculated.

The discriminant statistical model used for this analysis was:

$$\boldsymbol{DF = V_1X_1 + V_2X_2 + V_3X_3 + ... + V_IX_I,}$$

Where $\boldsymbol{DF}$ = discriminate function (score)

of grouping variables, $\boldsymbol{V}$ = the standardized discriminant coefficient or loadings for the clinical signs (predictors), $\boldsymbol{X}$ = respondent's score for the clinical signs, $\boldsymbol{I}$ = the number of predictor variables. The discriminant function coefficients V or standardized form beta indicate the partial contribution of each clinical signs to the discriminate function controlling for all other variables in the equation. They can be used to assess unique contribution to the discriminate function and therefore provide information on the relative importance of each variable. They also calculate the discriminant score for a given case. The score is calculated in the same manner as a predicted value from a linear regression. The magnitudes of these coefficients indicate how strongly the discriminating variables affect the score. They allow comparing variables measured on different scales. Coefficients with large absolute values correspond to variables with greater discriminating ability.

Results

In the present study, the initial presumptive diagnosis of both bovine babesiosis and anaplasmosis was achieved on the basis of case history, clinical symptoms, and microscopy, while confirmation was done using PCR assays and analysis of gene sequences. Six clinical variables including respiratory rate, hemoglobinuria, bloody feces, oculonasal discharge, mucous membrane, and size of lymph node were selected by DFA in first LDA, while three clinical finding including recumbent posture, nervous signs and oculonasal discharge were likely used for discrimination process in second LDA model (Table 1 and 2).

It was found that 89.0% of animals were correctly classified by using first LDA model. The average rate of correct classification was 78.8% (26/33) for *Anaplasma*; 100% (11/11) for *Babesia* infections (Table 3). On the other hand, the average rate of correct classification for survivors was 89.9% (134/149), whereas it was 57.1% (8/14) for non-survived subjects with 87.1% percent of correctly classified samples. The low accuracy of discrimination for non-survivors (57.1%) is due to not all dead animal showed abnormal posture or nervous signs so, these cases led to decrease the discrimination accuracy for non-survivors animals. The values of the statistical parameters F and Wilk's lambda for the discriminant functions showed a reasonable statistical significance. For the first LDA model, the first discriminant functions were identified to discriminate the type of infection. The function explained 94.2% of the variance with 0.058 Wilks' Lambda while the second one showed 62.4% of the variance with 0.376 Wilks' Lambda (Table 1). The first discriminant functions identified oculonasal discharge, bloody feces and hemoglobinuria (1.74, −0.894 and 0.576) as the important predictors for the first function to discriminate anaplasmosis from negative to TBDs, while bloody feces and respiratory rate were the best predictors for the second one for identifying the type of infection as these variables had large coefficients (−0.850, 0.728), respectively. Whereas score for other variables was less successful for prediction as these variables had small coefficients so it considered low predictors.

The first LDA model:

DF1 = − 0.282*respiratory rate
− 0.495*lymph node
− 0.434*mucous membrane
+ 1.735*oculonasal
+ 0.576*Hemoglobinuria
− 0.894*bloody faces,

DF2 = 0.728*respiratory rate
− 0.373*lymph node
+ 0.620*mucous membrane
− 0.252*oculonasal
+ 0.295*Hemoglobinuria
− 0.850*bloody faces.

Using this function, the discriminant scores for all cases were calculated. Cases with score near

Table 1. Results of two LDA models

Model	Wilks' Lambda	Chi-square	Degree of freedom	p-value
First LDA				
Function 1	0.058	309.361	12	<0.001
Function 2	0.376	105.988	5	<0.001
Second LDA	0.655	46.744	3	<0.001

Table 2. Loadings (standardized coefficients) and centroids in the discriminant function in two LDA models. Bold: important predictors for the discrimination

Clinical finding (Predictor)	Loadings (Coefficients)			
	First LDA Function 1	Function 2	Second LDA	
Oculonasal discharge	1.735	−0.252	0.558	
Hemoglobinuria	0.576	0.295		
Respiratory rate	−0.282	0.728		
Mucous Membrane color	−0.434	0.620		
Lymph node size	−0.495	−0.373		
Bloody feces	−0.894	−0.850		
Nervous signs			0.464	
Posture			0.510	
Centoroid				
First LDA	Function1	Function 2	Second LDA	
Anaplasma	6.550	1.490	Survived	2.201
Babesia	0.663	−1.956	Non survived	−0.235
Negative to TBDs	−1.342	0.688		

to the group centroid are predicted as belonging to that group, so any new cases being added can be classified (Table 2). For the second LDA model, one function was identified which explained 34.5% of the variance with 0.655 Wilks' Lambda. This function identified oculonasal discharge, recumbent posture and nervous sign (0.558, 0.510 and 0.464) as important or strong predictor to discriminate survivors from non survivors as these variables had large coefficients; whereas score of other variables were less successful for prediction as these variables had small coefficients, thus they considered low predictors. From this function we calculated the discriminant scores for all cases and compared with the group centroid (Table 2) as previously mentioned.

Second LDA model:

DF2 = 0.558*oculonasal discharge
+ 0.464* nervous signs
+ 0.510* recumbent posture.

Discussion

In the present study, it was appeared that 145 animal (89.0%) were correctly classified by using first LDA. The average rate of correct classification for *Anaplasma* and *Babesia* was 78.8 and 100%, respectively; while, the second LDA showed that 87.1% of the animals were correctly classified as survivors and non-survivors. The observed clinical signs were commonly

Table 3. Classification table for the two LDA models

First LDA			Prediction				Second LDA			Prediction		
			Anaplasma	*Babesia*	Negative to TBDs	Total				Not survived	survived	Total
Original	Count	*Anaplasma*	26	2	5	33	Original	Count	Not survived	8	6	14
		Babesia	0	11	0	11			Survived	15	134	149
		Negative to TBDs	11	0	108	119						
	%	*Anaplasma*	78.8	6.1	15.2	100.0		%	Not Survived	57.1	42.9	100.0
		Babesia	0	100.0	0	100.0			Survived	10.1	89.9	100.0
		Negative to TBDs	9.3	0	90.7	100.0						
Cross-validated[#]		*Anaplasma*	25	2	6	33	Cross-validated[#]		Not survived	5	9	14
		Babesia	0	11	0	11			Survived	15	134	149
		Negative to TBDs	12	0	107	119						

#: classification table of leave-one-out cross-validation

present with most subjects and the clinical examination of affected animals alone was not successfully revealed the underlying pathogen species. Even a combination of clinical signs and microscopic investigation did not substantially improve the diagnosis, because the clinical signs and symptoms were not identical and greatly confused with many other diseases. Among the tested clinical variables for the first model, the first discriminant functions identified oculonasal discharge, bloody feces and hemoglobinuria as the important predictors for the first function while bloody feces and respiratory were the best predictors for the second one for identifying the type of infection, respectively; while oculonasal discharge, recumbent posture and nervous sign were likely to prone complications and death. Some researchers have been recently reported that anemia and staring coat were important signs associated with the chronic syndrome of Babesiosis in cattle[12,17]. Hemoglobinuria, pale mucous membranes with empty episcleral blood vessels with reduced appetite have been attributed to severe hemolytic process associated the presence of piroplasms inside the red blood cells[5,6].

DA has been evaluated by few researchers for the discrimination of parasitic infections[3,6,7]. In one report, discriminant analysis was used to identify the most reliable markers of the diagnosis of ovine anaplasmosis and formulate the most appropriate prognostic variables of practical clinical importance[3]. Other investigators have used the forward DFA of the 4T samples and indicated that the test was significant in the classification of cod to eastern or western 4T[6]. While other researchers have used discriminant analysis to accurately distinguish between Diplectanum cazauxi and Diplectanum bauchotae[7]. DA can be useful for many types of objectives in parasitology and also veterinary researches.

Conclusion

The results herein demonstrated that discriminant analysis could provide information to increase the understanding of the discrimination among species of TBDs and could help predict a group membership for selection process as well as determining the best predictor for discrimination function. In this context, the first discriminant functions identified oculonasal discharge, bloody feces and hemoglobinuria as important predictors while bloody feces and respiratory rate were the strong predictors in the second function for discrimination of babesiosis and anaplasmosis, respectively; while nervous cattle with abnormal posture and oculonasal discharge likely to die. Future studies are needed for further confirmation of this predictor for discrimination and classification.

Acknowledgment

The authors would like to thank Dr. Mohamed El-Beskawy for his kind support in sample collection. The authors appreciate the great help of Professor Dr. Heinrich Neubauer, the director of the Friedrich Loeffler Institute (FLI), Federal Research Institute for Animal Health, Jena, Germany and Dr. Helmut Hotzel and Dr. Herbert Tomaso for their kind help and support for characterizing the samples.

References

1) El-Ashker M, Hotzel H, Gwida M, El-Beskawy M, Silaghi C, Tomaso H. Molecular biological identification of Babesia, Theileria, and Anaplasma species in cattle in Egypt using PCR assays, gene sequence analysis and a novel DNA microarray. Vet Parasitol 207, 329–34, 2015.
2) El-Ashker M, Salama M, El-Sebaei M, Risha E, Abdelhamid F, El-Diasty E, El-Fadl E. Significance of clinical variables and selected biochemical markers to predict the outcome of bovine anaplasmosis. Vet Med 60, 301–308, 2015.
3) Ciani E, Alloggio I, Petazzi F and Pieragostini E. Looking for prognosticators in ovine anaplasmosis: discriminant analysis of clinical and haematological parameters in lambs belonging to differently susceptible breeds experimentally infected with Anaplasmaovis. Acta Vet Scand 55–71, 2013.
4) Elsify A, Sivakumar T, Nayel M, Salama A, Elkhtam A, Rizk M, Mosaab O, Sultan K, Elsayed S, Igarashi I, Yokoyama N. An epidemiological survey of bovine Babesia and Theileria parasites in cattle, buffaloes, and sheep in Egypt. Parasitol Int 64, 79–85, 2015.
5) Fujinaga T. Bovine Babesiosis in Japan clinical and clinicopathological studies on cattle experimentally infected with Babesia ovate. Jpn J Vet Sci 43, 803–813, 1981.
6) Georgi J. Georgi ME and Theodrides VJ. Parasitology for Veterinarians, 5th ed. W. B. Saunders Company. Harcourt Brace Jovanovich, Inc. Philadelphia. 1990.
7) Giovanni S, Francesca B, Fabrizio S and Paolo G. Morphometric discriminant analysis for the classification of Diplectanum (Monogenea: Monopisthocotylea), parasites of Sphyraenaflavicauda. Parassitologia 47, 237–239, 2005.
8) Johnson RA and Wichern DW. Applied multivariate statistical analysis, 3rd ed. Englewood Cliffs Prentice-Hall. New Jersey, 1992.
9) Jonsson NN, Bock RE, Jorgensen WK. Productivity and health effects of anaplasmosis and babesiosis on Bosindicus cattle and their crosses, and the effects of differing intensity of tick control in Australia. Vet Parasitol 155, 1–9, 2008.
10) Joseph FHJR, William CB, Barry JB and Rolph EA. Multivariate Data Analysis, Seventh ed. Pearson Prentice Hall. 2010.
11) Kocan KM, Fuente JDL, Blouin EF, Coetzee JF, Ewing SA. Thenatural history of Anaplasmamarginale. Vet Parasitol 167, 95–107, 2010.
12) Losos GJ. Infectious Tropical Diseases of Domestic Animals, Longman Scientific & Technical Harlow CM20 2JE UK. 1986.
13) McClelland G, Melendy J. Use of parasites as tags in delineating stocks of Altaic cod (Gadusmorhua) from the southern Gulf of St. Lawrence and the Cape Breton Shelf. Fisheries Research 107, 233–238, 2011.
14) McLachlan GJ. Discriminant Analysis and Statistical Pattern Recognition. Wiley Interscience. USA, 2004.
15) Morrison DF. Multivariate Statistical Methods, 3rd ed. McGraw-Hill Publishing. New York, 1990.
16) Stockburger DS. Multivariate Statistics: Concepts, Models, and Applications. 2nd ed. Missouri State University, USA. 2001.
17) Urquhart GM, Armour J, Duncan JL, Dunn AM, Jennings FW. Veterinary Parasitological 2nd ed. Blackwell Science Ltd, Oxford. 1996.