EFFECT OF EXTRA BINOMIAL VARIATION ON PARAMETER ESTIMATES AND STANDARD ERRORS IN APPLICATION OF MULTIPLE LOGISTIC REGRESSION TO VETERINARY EPIDEMIOLOGY

Goelema[1], J.O., Jansen[2], J., Frankena[1], K.

Multiple logistic regression (LR) is often used to determine the effect of specific factors on disease occurrence. In veterinary epidemiology disease occurrence is usually measured at animal level, whereas most interest may be in farm level factors, among others herd production level, herd size, feeding strategy, pasturing conditions, hygiene measures etc.. Animal factors (i.e. parity, breed, stage of lactation) might act as confounders and one may wish to control for that.
Clustering of animals in herds or flocks leads to difficulties with the statistical analysis of binomial data because observations are no longer independent. Variation between herds or flocks is an extra source of variation and leads to extra binomial variation (EBV). Ignoring extra binomial variation may affect the significance of the factors (McDermott, 1988). In literature results of epidemiological studies are presented in which analysis of variables measured at cow level is carried out seperately from analysis of variables measured at farm level. However, variables measured at cow level and at farm level can be analyzed simultaneously, provided that the used model has the possibility to distinguish between variation between farms and variation between cows within farms. Jansen (1990, 1992) described an extension of the Logistic Regression model in which this distinction can be made. This extended model will be referred to as the Logistic Regression Random Effect model.

MATERIAL AND METHODS

The Logistic Regression Random Effect model (LRRE) is used to analyze farm level risk factors, while accounting for animal level variables for lameness in dairy cattle. Results of the model using LRRE are compared with results of multiple logistic regression (LR).

In the example data originating from a cross-sectional study on dairy farms in the province of Utrecht, The Netherlands, were used. Records of 2537 cows (parity>=1) (58 farms) were used for the analysis of Dermatitis Digitalis (Italian footrot). Effect of cow (parity, breed, stage of lactation, presence of other lameness symptoms) and environmental factors (herd size, herd production level, certain hygiene measures, cubicle size, cubicle type, condition of slatters, access to pasture, walking distance on metalled path and grassland soil type were evaluated using both LRRE and LR. Parameter estimates and their standard errors are compared.

[1]Dept.of Animal Husbandry, Animal Health section, Agricultural University, P.O.Box 338, 6700 AH Wageningen, The Netherlands

[2]Centre for Plant Breeding and Reproduction Research, Department of Population Biology, P.O.Box 16, 6700 AA Wageningen, The Netherlands.

Proceedings of the 6th International Symposium on Veterinary Epidemiology and Economics, 1991
Available at www.sciquest.org.nz

The data consist of binary observations Y_{ij} (1= diseased, 0= not diseased), where index i refers to farm i(=1 .. I) and index j(=1 ...J_i) to cow j on farm i. Given farm i, the probability p_{ij} that cow j on farm i is diseased, is given by

$$P_{ij}=\exp(\eta_{ij}+\sigma e_i)/1+\exp(\eta_{ij}+\sigma e_i) \quad ,$$

where $\{e_i\}$ are independent standard normal random variables; σe_i represents between-farm variation. In (1) n_{ij} will be related to K independent variables, whose values for cow j on farm i are arranged in the K*1 vector $\mathbf{x}_{ij}$, by means of a linear model $n_{ij} = \mathbf{x}^T_{ij}\beta$, where β is a K*1 vector of unknown regression parameters. Finally, conditional upon the farm contributions $\{\sigma e_i\}$, the observations Y_{ij} are independently distributed.

Model (1) will be referred to as a Logistic Regression Random Effect model (LRRE). The ordinary Logistic Regression model can be obtained by setting σ to zero. The above model can be fitted to data by the maximum likelihood method. Full details can be obtained from Jansen (1992).

All variables were included in the linear model without any screening in advance. Interactions were not tested for ease of interpretation (see table 1). After analysis of the full model, parameter estimates and corresponding standard errors were evaluated (see table 1). Relative changes were calculated to determine the difference between results of LR and LRRE. Significance of independent variables was considered by comparing the ratio of parameter estimates and standard errors with the <u>t</u> distribution.

The relative importance of the between-farm variation can be expressed by means of a coefficient of determination

$$\underline{R}=\{3\sigma^2/\pi^2\}/\{1+3\sigma^2/\pi^2\}.$$

The factor $\pi^2/3$ is the variance of the standard logistic distribution.

RESULTS AND DISCUSSION

Comparison of the standard errors presented in Table 1 shows a marked difference between LR and LRRE with respect to variables measured at farm level. By using LRRE standard errors of variables measured at the farm level showed average increases of 42.57% [range 0-100%]. For variables measured at the animal level the average increase found was 8.34% [range 2.33-9.91]. The estimate of σ yielded 1.071 (s.e. 0.1621), which means that the R is equal to 0.26.

Due to method of analysis, parameter estimates also showed changes . Decreases and increases alternated. Cow level parameter estimates changed by 13.95% [range 0.58-50.22], farm level variables changed by 46.10% [0-132.80] (all are absolute values).

The increase of standard errors and change of parameter estimates of the variables have a great impact on the significance of these variables. By using LRRE the significance of 11 variables changed. Ten of these were farm level variables. For one herd variable the estimate changed from positive to negative. Thus, for this variable the conclusions about the status changed from risk to preventive.

It is concluded that in situations where clustering in the data occurs, the statistical method used for analyzing such data should take into account the fact that observations within clusters are dependent. Ignoring the clustering may lead to drawing wrong conclusions about the significance of contributions of independent variables.

Statistical software that enables one to analyze firstly cow level variables and farm level variables simultaneously, and secondly take into account the extra variation caused by the clustering effect, therefore is highly recommended for analysis of possible risk factors for disease occurrence. It is therefore concluded that the Logistic Regression Random Effect model is a valuable tool in veterinary epidemiology.

REFERENCES

Jansen, J., 1990. On the Statistical Analysis of Ordinal Data when Extravariation is Present. Appl. Statist. 39, No.1:75-84

Jansen, J., 1992. Statistical analysis of threshold data from experiments with nested errors. Computational Statistics and Data Analysis (in press).

McDermott, J.J., Lesnick, T.G., and Martin, S.W., 1988. The analysis of individual animal risk for animals sampled in clusters. In: Acta Veterinaria Scandinavica, Proceedings of the 5th International Symposium on Veterinary Epidemiology and Economy: 459-461.

Curtis, C.R., Mauritsen, R.H., Salman, M.D. and Erb, H.N., 1988. The enigma of herd: a comparison of different models to account for group effects in multiple logistic regression analysis. In: Acta Veterinaria Scandinavia, Proceedings of the 5th International Symposium on Veterinary Epidmiology and Economy: 462-465.

Table 1. Parameter estimates (est.), standard errors (s.e.) and significance[1] of the independent variables from the analysis of Dermatitis Digitalis (DD) using both Logistic Regression (LR) and Logistic Regression Random Effect (LRRE).

		LR		LRRE		Change	
Variable	classes	est.	s.e.	est.	s.e.	est.	s.e.
Intercept		6.39	2.11*	4.46	3.03	30.2	43.2
Parity	Parity=1	0.92	0.17*	0.88	0.18*	4.0	3.9
	Parity=2	0.72	0.17*	0.65	0.18*	8.9	3.8
Lactational	Dry	0.23	0.18	0.11	0.19	50.2	6.6
stage (dim=	1 -<60 dim	0.47	0.26o	0.47	0.28*	0.6	5.9
days in milk)	60-<=120 dim	0.19	0.17	0.18	0.18	6.3	4.4
Breed(1)	>50% HF	0.21	0.16	0.12	0.17	43.1	5.9
	>50% FH	-0.28	0.20	-0.33	0.21	17.2	6.5
	>50% MRY	-1.04	0.42*	-1.08	0.46*	4.0	9.9
	Other	-2.08	0.74*	-1.94	0.76*	6.6	2.9
	50%HF/50%FH	reference					
Footrot grade>=2	Present	0.72	0.15*	0.70	0.15*	1.8	5.0
Sole ulcer	Present	-0.56	0.31o	-0.53	0.31o	6.1	2.3
Interdigital Hyperplasia	Present	1.82	0.19*	1.86	0.21*	2.3	8.3
Herd size	<=50 cows	-0.06	0.21	-0.08	0.28	45.4	34.8
	>=70 cows	-0.94	0.24*	-0.77	0.31*	17.6	27.7
Herd milk production	continuous (per 500 kg)	-0.00	0.00*	-0.00	0.00*	0.0	0.0
Cubicle cleaning	>2 times daily	0.16	0.22	0.09	0.34	43.4	57.3
Cows sleeping on slatters instead	Present of in cubicles	0.87	0.20*	0.43	0.30	50.5	48.5
Use of chalk	Yes	0.28	0.22	0.21	0.30	24.1	40.3
Standing bath	Present	-0.24	0.34	-0.04	0.46	84.8	36.9
Slot width	<40 mm	0.00	0.21	-0.58	0.35o	(23400)	61.8
Slat width	<126 mm	-0.22	0.17	-0.46	0.21*	115.2	24.2
Slatter surface	Rough	0.30	0.16o	0.54	0.25*	77.5	60.0
Irregular slatters	Present	-0.21	0.22	-0.37	0.29	74.8	29.4
Cubicle width	<1.10 m	-0.46	0.35	-0.13	0.56	72.5	58.1
	>1.10 m	-0.30	0.23	-0.44	0.33	45.9	45.1
Cubicle length	<2.20 m	0.11	0.27	0.26	0.49	132.8	81.3
	>2.20 m	0.55	0.19*	0.53	0.24*	3.4	20.9
Shoulderrail height	Continuous in cm	-0.05	0.02*	-0.04	0.02	18.3	41.8
Cubicle type	1(2)	0.47	0.19*	0.47	0.24o	0.2	23.9
	2	0.50	0.22*	0.42	0.29	14.8	27.9
Acces to pasture	Limited	0.78	0.46o	1.60	0.63*	104.9	36.6
Soil type	Peat	0.08	0.22	0.07	0.33	11.4	48.9
Walking distance on metalled path	Continuous in m	0.00	0.00	0.00	0.00	0.0	100.0
Hours outside	Continuous	0.08	0.04*	0.16	0.05*	107.8	30.4
Selection for better leg conformation	Yes	-0.01	0.20	-0.21	0.28	(1960)	37.8
Selection on claw pigment	Yes	-0.61	0.36o	-0.24	0.53	61.0	48.2
σ		-		1.07	0.16		

[1]*:P<0.05, o:p<0.10

(1):HF=Holstein Frisian, FH=Frisian Holstein, MRY= Meusse Rhine Yssel,
(2):Type 1:'R'-type fence, type 2: other, 'English' type= reference.

Proceedings of the 6th International Symposium on Veterinary Epidemiology and Economics, 1991
Available at www.sciquest.org.nz