PREDICTION OF CLINICAL DISEASE IN DANISH PIG HERDS

Baadsgaard NP[1], Grohn YT[2], Schukken YH,[2].McCulloch CE [3]
[1]Danish Institute of Agricultural Sciences, Research Centre Foulum, PO BOX 50, 8830 Tjele, Denmark;
[2]Department of Population Medicine and Diagnostic Science, Ithaca, and
[3]Department of Statistics, Cornell University, Ithaca, NY 14853, USA.

Throughout the clinical decision process the clinician will make use of his/her prior knowledge from previous observations, and although often used implicitly by most experienced clinicians, the formulation of prior knowledge or a priori expectation is an important starting point in clinical work[1]. The objective of this study was to develop a predictive model for major clinical findings in pigs based on repeated data collections from 8 Danish pig herds. The methods that were investigated were: ordinary logistic regression, generalized estimating equations and nonlinear mixed models. The results from clinical observations on findings related to respiratory disease and poor body condition are compared and interpreted as well as intra-class correlation for 7 major clinical findings.

Materials & Methods

The data for this study came from 8 pig herds with conventional health status and a continuos production system. The size of the herds varied from 200 to 300 sows. The local veterinarians visited the herds on a monthly basis. At each visit they carried out a systematic clinical examination in a fixed representative sample of pigs in the age interval from 0-40 kg. The average sample size was 60. The clinical examination consisted of binary recordings of 7 major clinical findings: poor body condition (PBC), dullness, skin diseases, and diseases of the locomotive, central nervous, digestive, and respiratory systems.
The interest was in comparing different methods for estimating the correlation within and among these diseases over time. The recordings were made at the flock level. Hence, the unit of observation was the individual herd. To allow for any seasonal variation, season was included, together with potentially and biologically meaningful correlated clinical findings, as predictors in the model. The regression model was a logistic regression model with a binomial error distribution. The three different estimation procedures that were considered were the following:

1. Logistic regression model, ignoring correlation within herds. Herd treated as fixed effect.
2. Generalized Estimating Equation (GEE) allowing for within herd correlation with no herd specific parameters included.

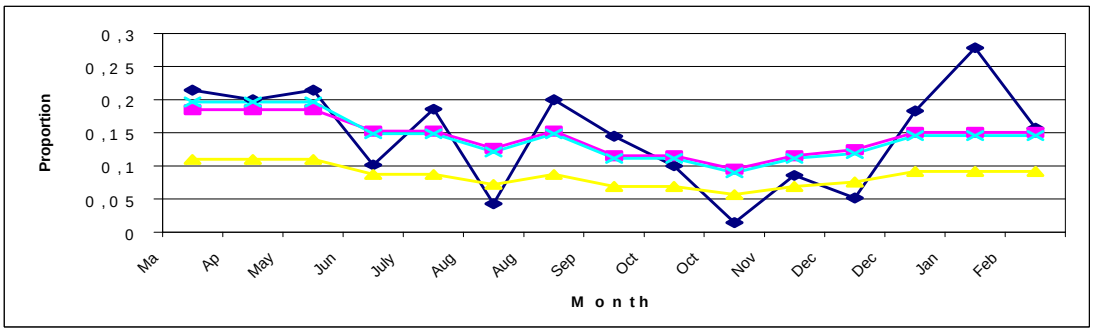
3. Nonlinear mixed model with fixed and random effects, i.e., herd was allowed to have a nonlinear relationship to the response variable[2].

Consider the marginal expectation of binary outcome Y_{it}, where *i* indexes clusters, i.e., herds, and *t* denote units within clusters. In Model 2 the outcome was modeled as a function of the covariates but with no herd specific parameters. In the nonlinear mixed model the probability distribution of Y_{it} was modeled as a function of the covariates and a random parameter specific to the *i*'th cluster. All analysis was done in SAS v7.0. Models 1 and 2 were run in PROC GENMOD. Model 3 was run in PROC NLMIXED.

Results

The predicted values for Models 1 and 3 fit the observed values well while Model 2 performs poorly in making herd specific predictions (Fig. 1). The predicted values for this model were centered around the mean value for poor body condition for all herds.

Figure 2. Plot of observed porportions of pigs from one herd with respiratory disease versus predicted from the logistic regression, the genmod procedure and the nlmixed procedure. Legend: ◆=observed, ■=logistic regression, ▲=genmod procedure, ✕=nlmixed



Digestive disease was a significant predictor for poor body condition (Table 1). Parameter estimates and standard errors increase from a simple model to a model in which the variance term is included.

Table 1. Parameter estimates and standard errors for poor body condition and respiratory disease for the three estimation procedures. * Prediction error~(observed-predicted)[2]

Parameters	Proc logistic	Proc genmod	Proc nlmixed
Model:			
Poor body condition			
Prediction error	0.53	1.13	0.55
Season	Non Sign	Non Sign	Non Sign
Herd	Sign		0.133
Digestive disease(std error)	1.63(0.39)	1.69(0.40)	1.70(0.49)
-2log likelihood	6608	6770	610
Model:	Proc logistic	Proc genmod	Proc nlmixed
Respiratory disease			
Prediction error	0.20	0.41	0.20
Season	Sign	Sign	Sign
Herd	Sign		0.469
-2log likelihood	3088	3270	489

Correlated clinical diseases were included as predictors for 5 clinical findings (Table 2). The results indicated only low correlation within herd; the intra class correlation coefficient was less than 0.1. For respiratory disease there was a moderate correlation within herds; the intra class correlation coefficient was 0.124.

Table 2.Correlated diseases, variance components and intraclass correlation coefficients from nlmixed

Disease	Correlated disease	Herd variance component	Standard error	Intraclass correlation[a]
Poor body condition	Digestive disease	0.133	0.066	0.038
Alertness	Poor BC	0.186	0.098	0.053
Skin disease	Leg disease	0.301	0.145	0.083
Leg disease	Alertness	0.269	0.147	0.075
CNS disease	None	0.299	0.191	0.083
Digestive disease	Poor BC	0.187	0.097	0.053
Respiratory disease	None	0.469	0.239	0.124

[a]intra class correlation was calculated as: Var. comp./(Var. comp.+3.3)[3]

Discussion

Weak correlation within disease was observed. Information on the spatial clustering among diseases contributed significantly to the models for all diseases except respiratory disease and central nervous disease. Models that allowed random variation tended to yield slightly higher values for point estimates and larger standard errors while there was no difference in the predicted values. The validation of a predictive model should be assessed not only by data fit, but also by evaluating the ability of the model to make forecasts based on subsets of the data set.

References:

[1]Feinstein, A.R. An Analysis of Diagnostic Reasoning I-III The domains and Disorders of Clinical Macrobiology. 1973;Yale Journal of Biology and Medicine 46: 212-232.

[2]Pinheiro, J. C., Bates, D.M. Approximations to the log likelihood function in the nonlinear mixed effects model. 1995; Journal of Computational and Graphical Statistics 4: 12-35.

[3]Atwill, E.R., Mohammed, H.O., Lopez, J.W., McCulloch, C.E., and Dubovi, E.J Cross-sectional evaluation of environmental, host, and management factors associated with the risk of seropositivity to *Ehrlichia risticii* in horses of New York State. 1996; American Journal of Veterinary Research, 57: 278-285.