

BAYESIAN NETS FOR MASTITIS DIAGNOSIS IN DANISH ORGANIC AND CONVENTIONAL DAIRY HERDS

VAARST, M.^a, THYSEN, I.^b, ENEVOLDSEN, C.^a, CHRIÉL, M.^c, THIESSON, B.^d,
ANDERSEN, S.K.^d, AGGER, J.F.^c

Bovine mastitis is a complex multifactorial disease with multivariate manifestations. Because of this complexity, expert system methodology and statistical models may have the potential to synthesise the available data from dairy herds and, consequently, improve the diagnostic efficiency in mastitis control. The purpose of this study was to 1) develop expert systems based on Bayesian nets (BN) for mastitis pathogen diagnosis and 2) evaluate whether the developed models behaved consistently across markedly different dairy herds, thus allowing general application of the models.

MATERIALS AND METHODS

Bayesian nets (Hogeveen et al., 1994) were developed from expert knowledge and data from conventional and organic dairy herds. Certified organic dairy production differs from conventional production in the following respects: medical disease prevention is prohibited, withdrawal time after medical treatment is three times prolonged, and straw bedding and daily exercise must be provided to all animals. In general, farmers of organic herds treat non-severe cases of clinical mastitis by handmilking and/or hyperaemia-inducing ointments.

Data were collected from 367 mastitis cases from 14 organic herds and 487 mastitis cases from 34 conventional herds. Proportions treated by a veterinarian were 26% and 100%, respectively. Milk was sampled aseptically from mastitic quarters; in organic herds by the farmer immediately after having recognised abnormal clinical changes in milk, udder or teat; in conventional herds by the local veterinarian before treatment. The general condition of the cow and local symptoms were recorded by the sampler. The milk samples were examined bacteriologically for presence of pathogens.

Two different approaches were applied to build an initial BN from the data collected in the organic herds:

1) '*Manual approach*': First, a multiple correspondence analysis (MCA) was performed to identify dimensions in the data (data reduction). Second, logistic regression (LR) was used to estimate direction and magnitude of associations indicated by the MCA. Based on the LR-estimates a network structure ('path diagram') was suggested.

2) '*Semi-automatic approach*': First, the variables were divided into a response structure of groups which reflect direction of relations, i.e. predictor variables precede the 'disease variable' (pathogens isolated at incident), which again precede symptoms. Second, the BIFROST software (Højsgaard and Thiesson, 1994; Højsgaard, et al., 1992) was applied to these groups to automatically create a network structure from the observed data by a statistical model selection procedure. BIFROST exploits the specified direction of relations between groups of variables and selects a model which obeys these relations.

^aDepartment of Research in Cattle and Sheep, NIAS, P.O. Box 39, DK-8800 Tjele, Denmark.

^bDina Foulum, P.O. Box 23, DK-8800 Tjele, Denmark.

^cDepartment of Animal Science and Animal Health, KVL, DK-1870 Frederiksberg, Denmark.

^dDepartment of Mathematics and Computer Science, AUC, DK-9220 Aalborg., Denmark.

The results of these two approaches were then combined with expert knowledge to create a final BN. Finally, the pEM software (Thiesson, 1991; Lauritzen, 1994) was applied to estimate the conditional probabilities to quantify the structural relations. Forty two mastitis cases from the organic herds were used to test the diagnostic efficiency of the final BN (the 'expert/pEM model'). These cases were not used for model selection and parameter estimation. Finally, the 'organic BN' was applied to the 'conventional' data to evaluate the diagnostic efficiency.

RESULTS AND DISCUSSION

The results of the manual and the semi-automatic approaches to develop network structures were rather consistent. The interaction between the manual and the semi-automatic model building approach was a very fruitful process, and the study has indicated good prospects for replacing the often tedious manual approach with an iterative semi-automatic selection approach which exploits software like BIFROST.

The final BN (Fig. 1) contains two latent (unobservable) variables describing the possible chronic and acute infections, both by four mastitis pathogens (*Staph. aureus*, *Strep. dysgalactiae*, *Strep. uberis*, and coliforms). The BN furthermore contains two predisposing variables and five variables expressing characteristic symptoms. Based on expert knowledge and knowledge represented by data, the pEM software was used for the quantification of the structural relations in the model. Expert knowledge was used to provide *a priori* knowledge for conditional probabilities on the effects of the predisposing factors, on the relation between a chronic pathogen and the pathogen causing the acute clinical symptoms (assumed with high probability to be the same pathogen), and of the relation between the actual presence of a pathogen and the bacteriological findings (assuming 50% sensitivity of identifying *Staph. aureus* and coliforms, and 95% sensitivity of identifying the two other pathogens). The relations between the disease variables and the symptoms were not provided with expert *a priori* knowledge.

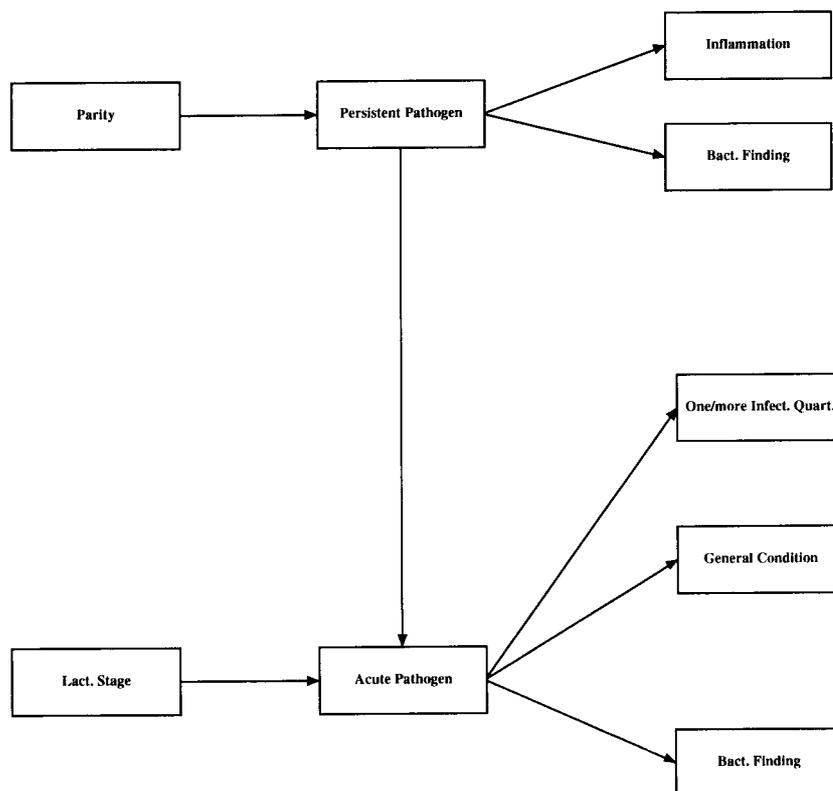


Fig. 1 The final Bayesian Net for diagnosing mastitis pathogens

When applied to the validation subset (42 cases), the BN could only diagnose a pathogen with at least 50% certainty in 18 cases. Of those, 15 were correctly classified. The three errors were *Strep. uberis* misclassified as coliforms. Overall, this analysis indicated a low sensitivity but a relatively high specificity. If the prior bacteriological diagnoses were not used, both sensitivity and specificity were low. Mastitis symptoms in general had a very poor predictive ability.

The application of the organic BN to the conventional data (which had no information of prior bacteriological diagnoses) also showed a very poor sensitivity and specificity. In this case the model, in particular, misclassified other pathogens as being *Staph. aureus*. This test indicates that it is unlikely that a model will work consistently across different production systems. Differences in sampling procedures and different attitudes towards mastitis control and mastitis treatment routines in the herds may be major explanations for the poor external validity of the organic BN model. Sampling in organic herds was performed at a very early stage of the mastitis process, where the clinical manifestations are expected to be less specific. Consequently, future research should focus on developing adaptive prediction models which can 'learn from' the specific environment in which they are applied. Ideally, herd specific models which take into account dynamic aspects related to the herd, as well as the cow, should be developed.

Bayesian networks are well suited for model-based decision support systems (DSS) in domains characterised by uncertainty. A further advantage is the availability of efficient statistical methods to combine expert knowledge and data analysis in the model development. A DSS for diagnosing mastitis pathogens has potential practical benefits for choosing between medical and alternative treatments of clinical mastitis, identification of individual animals with chronic subclinical mastitis for segregation or treatment, and cost reductions by reducing the number of laboratory examinations of milk samples. To obtain full benefits of such a DSS it must be integrated in some ongoing herd health and productivity recording system.

REFERENCES

- Hogeveen, H., Thiesson, B., Andersen, S.K., Schukken, S.K. and Noordhuizen-Stassen, E.N., 1994. Approaches to the building of Bayesian Belief Networks for disease diagnosis. Proc. 7th International Symposium on Veterinary Epidemiology and Economics, Proc. 7th International Symposium on Veterinary Epidemiology and Economics, Nairobi, Kenya, 1994.
- Højsgaard, S. and Thiesson, B., 1994. BIFROST - block recursive models induced from relevant knowledge, observations, and statistical techniques. Comput. Statist. & Data Analysis (in press).
- Højsgaard, S., Skjøth, F., and Thiesson, B., 1992. User's Guide to BIFROST, Ver.1.0, Rep. R-92-2001, Dep. of Mathematics and Computer Science, Aalborg University, Denmark
- Lauritzen, S.L., 1994. The EM algorithm for graphical association models with missing data. Computational Statistics & Data Analysis (in press)
- Thiesson, B., 1991. (G)EM algorithms for recursive graphical association models with missing data. Master's thesis, Institute for Electronic Systems, Aalborg University.