Epidémiol. santé anim., 1997, 31-32

EPIDEMIOLOGICAL CONSIDERATIONS RELATED TO WITHIN HERD MULTIVARIABLE MODELLING IN HERD HEALTH MANAGEMENT

Enevoldsen C.[1]

Les modèles multivariables ont été développés et utilisés en collaboration avec plusieurs vétérinaires praticiens spécialisés dans l'élevage bovin laitier danois. Le cadre de la modélisation est présenté ici, ainsi que des considérations épidémiologiques liées à la spécification des modèles et à l'interprétation des estimations des paramètres qui semblent être particulièrement importantes dans ce contexte. Malgré un large accès aux enregistrements de données concernant le traitement des maladies, le manque de données sanitaires cliniques fiables a été le principal obstacle pour une analyse utile. Ce déficit de données fiables a rendu les comparaisons des statuts sanitaires entre troupeaux difficiles. Un cadre avec de nombreux modèles spécifiques de parité relativement simples a été retenu pour aborder le problème de l'interaction et plusieurs séries de règles de décisions ont été élaborées pour détecter et éliminer les estimations de paramètres non pertinentes. Les estimations n'ont habituellement pas pu être considérées comme causales. Un recoupement avec des données qualitatives obtenues lors de visites de troupeaux a été généralement nécessaire pour une analyse causale.

Les analyses statistiques ont fourni une plate-forme pour caractériser les procédures de gestion et d'enregistrement. Le problème des comparaisons multiples a été reconnu mais traité par une approche qualitative. Une autre série de règles de décision a été développée pour identifier les structures de données importantes et pour sélectionner automatiquement ces relations pour l'écriture de rapport en tant qu'outil fournissant une entrée aux systèmes d'aide à la décision.

L'étude montre que le suivi (épidémiologie descriptive) et l'analyse s'imbriquent de plus en plus. Le besoin de meilleurs outils pour synthétiser un grand nombre d'estimations de paramètres a été mis en évidence. Des modèles avec des variables à réponse multiple devraient décrire de manière plus fiable le complexe santé-production.

THE PROBLEM

The ultimate goal of health and production management of livestock herds is to improve decisions and actions that promote health. The decision maker or the advisor applie some set of decision rules to the available information before making decisions or giving advice, respectively. In many herds most of these rules probably are implicit. Especially in cases of complex problems like when and how to treat sick animals or change a feed ration. Explicit rules usually are easier to identify in herds in case of simple problems (e.g. rules for selecting animals for preventive treatment). In recent years advanced decision support systems (DSS) have become available. Several of these DSS are very complex (simulation) models that require large numbers of input parameters that characterize the animals (production, reproduction and health), the production system (e.g., housing capacity, feed supply, and milk quota), and the management strategy (e.g., feeding strategy, culling rules, and voluntary wait period) (e.g., Enevoldsen et al., 1995).

Most health problems in livestock herds are inter-linked and these complex relations are usually influenced by numerous factors (risk factors or causes) which usually also are strongly inter-linked. Consequently, it is no simple task to characterize a health (or production) problem in a herd and identify those factors that influence the problem and the options for controlling them.

Often input parameters for decision making in a particular herd are derived from scientific studies or from experience developed more or less systematically from trial and error either in the herd in question or in other herds. Application of parameters from external sources (scientific studies or experience) may be highly problematic because there are numerous examples of inconsistent reports of relations between risk factors and occurrence of health problems. In addition, inconsistent results may be under-reported because they make less sense or are "non-significant" (publication bias). One simple explanation of this heterogeneity of effects is that there often is a true "biological" interaction between risk factor and herd (Enevoldsen, 1993). Such interaction should, in fact, also be expected if the theories related to necessary and sufficient causes of multifactorial diseases are accepted. That is, most relations between risk factors and problems may be unique in each herd. This heterogeneity of effects was clearly demonstrated by Markusfeld-Nir (1997) who found highly variable associations between 7 risk factors and a parity-specific measure of fertility in 141 Kibbutz dairy herds.

Consequently, input parameters to DSS should ideally be derived from the herd where the decisions will be made. Because most health problems are multifactorial, some sort of multivariable analysis must be conducted to provide such herd specific parameters. The purpose of this paper was to discuss the options and barriers related to such within herd multivariable epidemiological modeling with particular emphasis on the analytical problems related to smaller dairy herds (50-100 cows).

[1] Tornager 2, DK-7600, Denmark. E-mail: carsten.enevoldsen@post3.tele.dk

THE STUDY CONTEXT

A collaboration was established with 4 practicing veterinarians. They provided production, reproduction, and health data from some of the dairy cow herds in their practices. Typical herd sizes were from 50 to 200 cows. Linear and logistic regression (PROC GENMOD, SAS 6.11, 1996) was used to analyze a series of potentially important epidemiological questions. The choice of problem complexes for analysis (response and explanatory variables) was partly based on the models applied in the integrated Israeli health and production analysis system (Markusfeld-Nir, 1996), partly based on previous detailed herd health management studies performed by the author (Enevoldsen, 1996). In general, separate models were specified for each parity because most associations were expected to differ from parity to parity in each herd.

Results of analyses and model characteristics (e.g., fit and robustness) were evaluated by the author and a preliminary interpretation of the results was presented to the practicing veterinarian who used the results during herd health visits. The feed-back from these visits and the subsequent discussions with the veterinarians were used to modify the analyses. Further input to the development of the concept was obtained from presentations of the concept at workshops and informal meetings with practicing veterinarians. Finally, through continuous inspiration from the Israeli system.

DATA QUALITY CONSIDERATIONS

Recording of clinical data (e.g. lameness and body condition) was very different from veterinarian to veterinarian and from farmer to farmer and good agreement was extremely difficult to obtain (Baadsgaard & Enevoldsen, 1997). Recording was also often quite variable within person or within herd due to housing and management factors affecting the recording procedure. Farmers also often applied different threshold for spending treatment costs depending on production level and expectations to performance of the individual cows. Recording accuracy was also quite variable. Within person estimates were therefore preferable and between herd comparisons of health status often were not justifiable. Insufficient data quality clearly was the major analytical problem. Recordings of disease treatments were generally insufficient health information.

MODEL SPECIFICATIONS

Because of the relatively small herd sizes, models had to be simple and very few interaction terms could be included. The validity problems caused by this simplification are currently being assessed. Simplicity and interpretability of dependent and independent variables were other very important issues. It appeared that specification of a relatively large number of models within parity and a few stage of lactation groups was an acceptable solution to these problems although the risk of observing spurious associations may increase with this approach.

The following response variables are currently included in the analytical framework: 1) peak yield, 2) total milk in 305 days, 3) daily milk at drying off, 4) dry period, 5) not served at 150 DIM, 6) pre-service unobserved heat, 7) not pregnant at first service, 8) not pregnant at 150 DIM, 9) SCC 9-60 DIM, 10) SCC in late lactation, 11) culling 0-50 DIM, 12) culling 51-150 DIM, 13) culling after 150 DIM, 14) milk fever, 15) retained placenta, 16) udder treatment -21 to 50 DIM, and 17) udder treatment after 150 DIM. All responses except the culling variables and milk fever were analyzed separately for lactations 1, 2, and later.

The explanatory variables included in one or more model currently are the following: 1) summer calving, 2) reproduction disorder at calving, 3) digestive or metabolic disorder before 50 DIM and 4) in the entire lactation, 5) udder treatment before 50 DIM and 6) in the entire lactation, 7) high SCC in early lactation and 8) in the entire lactation, 9) dry period below 50 days, 10) low milk yield at drying off, 11) high milk yield at drying off, 12) high peak yield (primarily as an estimate of negative energy balance), 13) low peak yield (to allow for a non-linear relation), 14) drying off in November to February (to detect quota adjustments), 15) short voluntary wait period, 16) pre-service unobserved heat, and 17) parity (included in the analysis of culling profile and milk fever). High, low, and short was defined as 33 or 67 percentiles as appropriate. Which risk factors to use as predictors of each response variable was determined by an overall model structure that determines the sequence of events. Calving disorders, for instance, were included in the model of peak yield. Peak yield was included in the model of total 305 day yield but calving disorders were not because they were expected to primarily influence peak yield. This approach was expected to simplify the models and reduce multi-collinearity problems

ESTIMATION PROBLEMS

In smaller herds and with few counts in a category of a discrete response variable, model parameters often could not be estimated or model validity was questionable. A series of decision rules were included in the analysis procedures to delete clearly invalid parameter estimates and print warnings in case of questionable results. These rules differed from categorical to continuous responses.

MODEL INTERPRETATION

The study has clearly demonstrated that parameter estimates often reflected selective handling of cows and changes in management and recording. Milk yield, for instance, often was higher among cows with records of prior mastitis treatment. The veterinarians first hand knowledge about the herds could reveal that this simply was part of the culling strategy where the farmer only wanted to spend treatment costs on the high yielding cows he wanted to keep. Similarly, cows were often not inseminated because of high somatic cell counts. Consequently, the resulting parameter estimates rarely could be regarded as causal. The main result of the statistical part of the analysis is that it identifies structures in the data that deviate from expected. Such structures are valuable aids for getting a detailed insight into management policies and recording procedures in the individual herd. Because management is regarded as the most important determinant of health and productivity, this was a significant outcome. Such an insight probably is an efficient tool for more in-depth qualitative investigations of the complex management of the highly dynamic livestock herd systems. The causal analysis, therefore, comes after the statistical analysis and is based on a combination of quantitative and qualitative information. Because the framework includes a large number of statistical tests, a multiple comparisons problem is evident. Currently this problem is addressed qualitatively by developing rules for interpretation. For instance, requiring consistency of estimates across models and parities. This approach is similar to "pattern diagnosis" (Morris, 1982) and "the point of view concept" (Ducrot et al., 1996). This approach also means that monitoring (descriptive epidemiology) and analysis overlap more and more.

PRACTICAL CONSIDERATIONS

Because of the large number of parameter estimates, reading and interpreting the analysis reports could be quite time consuming. A set of rules was therefore included in the framework to select parameters for output to a main report including "significant" findings only. Model parameters were used to estimate both contributions of response variability of all factors together and each risk factor individually. Another result of the analysis is that a "standard cow" can be constructed. That is, the expected response from a cow (e.g., peak yield or not pregnant at 150 DIM) given she is not exposed to any risk factor. Such a standard cow will be useful for performing unbiased monitoring of the herd level responses.

NEEDS FOR FURTHER DEVELOPMENT

Further sophistication of the statistical models apparently is not needed currently because the lack of reliable data is the major analytical barrier. There is more need for more veterinarians using relatively simple multivariable models in this fashion in a larger number of herds and a subsequent systematic analysis of the results of this process, both quantitatively and qualitatively. However, the study has revealed a considerable need for more efficient tools to synthesize the large number of estimates from the applied analysis strategy. Some hierarchical model type might be useful. Also for more work on the qualitative approaches to causal analysis. Several processes will also be modeled more appropriately with a multivariate model (several response variables). This could be accomplished by using Bayesian Nets or through a factor analytic approach where "dairy cow types" could be identified and subsequently related to sets of risk factors.

BIBLIOGRAPHY

Baadsgaard N.P., Enevoldsen C. 1997. A potential approach to support animal welfare promotion in a Danish veterinary practice context. Pages 108-119 in Proceedings of the meeting 9th-11th of April, 1997 of the Society for Veterinary Epidemiology and Preventive Medicine. Liverpool, UK. Ed. MV Thrusfield & E.A. Goodall.

Ducrot C., Legay J.M., Gröhn Y.T., Enevoldsen C., Calavas D. 1996. Approach of complexity in veterinary epidemiology: Example of cattle reproduction. Natures-Sciences-Sociétés (4), 23-35.

Enevoldsen C. 1993. Sundhedsstyring i mælkeproduktionen (Dairy herd health management). PhD Thesis (in Danish). The Royal Veterinary and Agricultural University, Copenhagen, Denmark. 138 pp.

Enevoldsen C. 1996. Epidemiology in veterinary practice revisited. Pages 109-116 in Proceedings of the meeting 27th-29th of March, 1996 of the Society for Veterinary Epidemiology and Preventive Medicine. Glasgow, UK. Ed. MV Thrusfield & E.A. Goodall.

Enevoldsen C., Sørensen JT., Thysen I., Guard C., Gröhn YT. 1995. A diagnostic and prognostic tool for epidemiologic and economic analysis of dairy herd health management. J. Dairy Sci (78), 947-961.

Markusfeld (Nir) O. 1996. Integrated herd health programs - the Israeli experience. Pages 126-135 in Proceedings of the meeting 27th-29th of March, 1996 of the Society for Veterinary Epidemiology and Preventive Medicine. Glasgow, UK. Ed. M.V. Thrusfield & E.A. Goodall.

Markusfeld-Nir O. 1997. Body condition score in production and health management in dairy cows in Israel. Pages II.7.1 to II.7.12 in Proceedings of the meeting 17th-18th of April of Danish Buiatrics Society, Brædstrup, Denmark.

Morris RS.1982. New techniques in veterinary epidemiology - providing workable answers to complex problems. Pages 1-16 in Proceedings of the meeting 22nd-25th of September, 1982 of the Soc. for Veterinary Epidemiology and Preventive Medicine. Reading, UK.

SAS User's Guide: Statistics, Version 6.11 Edition, 1996. SAS Inst., Inc., Cary, NC, USA.