IDENTIFICATION OF INTERACTIONS AND NON-LINEAR EFFECTS, ASSESSMENT OF HERD EFFECTS, AND MODEL VALIDATION IN A DAIRY HERD HEALTH MANAGEMENT CONTEXT

Enevoldsen C[1], Elgaard KB[2], Jensen NP[3], Nielsen KH[4], Nørregaard HK[5], Philipsen JS[6]

[1]EPIVETKO, Tornager 2, DK-7600 Struer, Denmark. E-mail: carsten@epivetko.dk
[2]Midtsalling Dyrlæger, Helsevænget 2, DK-7870 Roslev, Denmark.
[3]Uggerslev Dyreklinik, Gl. Skolevej 36, DK-5450 Otterup, Denmark.
[4]Skovsgaard Dyreklinik, Præstegaardsvej 12, DK-9460 Brovst, Denmark.
[5]Tinglev Dyrehospital, DK-6360 Tinglev, Denmark.
[6]Skægkær Dyreklinik, Skægkærvænget 2, DK-8600 Silkeborg, Denmark.

The problems: The biological real-life problems addressed by epidemiologist are usually complex. Consequently, a considerable number of (explanatory) variables must be included in the applied statistical models to describe such problems. Due to the large number of variables, the number of possible (statistical and biological) interactions among variables is also large. Often non-linear relations are also highly plausible. Specification and selection of valid statistical models are, consequently, difficult tasks. It is common practice to reduce the number of explanatory variables through bivariate (one response and one explanatory variable) analyses prior to the multivariable analyses (one response and multiple explanatory variable) or by means of multivariate methods (multiple response variables) like factor analysis that produces scores to replace the original (explanatory) variables[1]. Both the bivariate and the multivariate approaches will almost certainly cause considerable loss of potentially important information. Although creation of scores or indexes by mean of factor analysis and related techniques are statistically attractive, the approach make biological interpretation of the results difficult. Epidemiological analyses are frequently conducted across herds to increase statistical power of significance tests and generality of the (practical) inferences but such analyses must handle the non-independence of observations within herd or other "clusters" like cow and parity correctly. Models to handle clustering have attracted much attention in recent years (e.g., random effects (mixed regression) models). Usually the applied approaches to handling herd effects neglect the possible occurrence of interaction between variables at different levels (e.g., cow and herd) although occurrence of such effects are highly plausible (e.g., culling of individual cows may be dramatically influenced by the milk quota). For these reasons results of traditional across-herd analyses may very well be (completely?) invalid if they are applied as recommendations to the individual farmer. The magnitude of this problem is indicated by a summary of within-herd estimation of the cow-level associations between some major diseases and milk production in 156 Israeli Kibbutz herds. This study clearly showed that the associations were highly variable from herd to herd[2]. The objective: An obvious solution to all these problems of across-herd analyses is to conduct separate within-herd analyses and use the results of these analyses to specify more appropriate across-herd models or combine the herd-level results systematically by means of a meta-analysis. Since early 1997 our group of one epidemiologist and five specialized practicing cattle veterinarians has conducted systematic multivariable within-herd analyses of data from Danish dairy herds serviced by the authors on a commercial basis. The purpose of this summary is to describe major experiences from this process. The material and metods: The framework of the multivariable analyses is described in detail elsewhere[3,4] and it is very similar to the nationwide Israeli system[2]. Models are either linear or logistic regression models – both ordinary least square and multilevel (mixed) models are applied. The response variables of primary interest are peak yield (average daily yield 9-92 days p.p.), projected daily yield at 305 days, projected 305-day yield (linear models), non-pregnant to first service, open at 150 DIM, culled before 50 DIM, and culled after 150 DIM (logistic models). Test-day milk yield is the response variable in the mixed models. Except for the culling and test-day models, analyses are conducted within parity (and within herd). A series of explanatory variables relevant to the response variables and the parity of interest are explored. Examples are calving season, age at calving, health status, and dry period length. Supplementary variables are becoming available from an increasing number of herds, which now also supply observations from the veterinarians' regular weekly clinical examinations of cows (mainly vaginal discharge and ketone reactions 1-2 weeks p.p., udder condition, and body condition score (BCS) at drying off, calving, and 6-12 weeks p.p.). In April 2000 approximately 26,000 recording dates (cow level information) are available from approximately 75 herds with weekly examinations of cows. The results of these analyses are used intensively for herd management purposes by the veterinarians. For instance, cows with positive ketone reactions will usually receive some sort of treatment and cows with very long dry periods or dramatic loss of body condition will be given special attention at calving. This process can be regarded as model validation although not conducted systematically because the entire process is conducted without any sort of external funding. Results and discussion - SCC: This process has clearly shown that estimates differ markedly from herd to herd and that parity-specific within-herd estimates (models) usually are needed. E.g., in some herds strong associations between SCC and milk yield were revealed while no associations were detected in others. In some herds, associations between milk yield and SCC were evident in all parities while only older cows were affected in other herds (the most "typical" pattern). Sometimes such heterogeneity of effects could be "explained" by the additional "non-recorded" information veterinarians supplied by the veterinarians (e.g., culture results and details about clinical manifestations). Results and discussion – age at first calving and BCS: Associations between age at first calving and milk yield were also highly variable from herd to herd. In some herds curvi-linear associations were detected, while straight-line relations were sufficient to describe relations in others. Consequently, the calving age associated with maximum milk yield was highly variable from herd to herd. Finally, an association between age at first calving and milk yield in second lactation was detected in some herds. More detailed measurements of body weight and size in some herds often provided plausible explanations to these between-herd differences. Most recently analyses of body condition scores indicate that interactions between the length of the dry period and lactation curve characteristics in the previous lactation probably must be taken into account to obtain valid models of early lactation milk yield. Results and discussion – disease treatment data: Numerous epidemiological studies have utilized records of disease treatments as indicators of health status. Our experiences indicate that this is very problematic in many herds because several farmers select cows for treatment according to expected future value of the individual cow. This "preferential" treatment is usually based on milk production potential. That is, cow with high production potential are more likely to receive treatment in case of disease signs like clots in the milk or the analysis indicates that "mastitis" is associated with higher production! The analytical problems related to such preferential treatments is well known with respect to breeding values and BST[5]. Some part of the strong genetic correlation between mastitis and milk yield reported in numerous studies may very well be due to this phenomenon. Similar problems arise with respect to interpretation of associations between clinical recordings and milk yield because treatment or preventive action according to expected effects of clinical observations probably will and should "hide" the biological effects of certain diseases. E.g., propylene glycol treatment to cows in excessive body condition at drying off or cows with positive ketone body reaction post partum. Often the intensity of the action taken will depend on some expectation to the outcome and these expectations are usually unknown to the analyst. In general such observational data rarely will allow estimation of "effects" but they probably will be useful to estimate whether potentially negative effects are reduced to negligible by the management routines introduced during recording of data. Results and discussion – test-day models: Prototype versions of multilevel test-day models of milk yield have been developed to assess lactation curve characteristics and seasonal influences simultaneously. Currently a three-parameter lactation curve model appears to fit the data satisfactorily (peak yield fixed at 60 days p.p.). By centering DIM at 60 days (intercept) and re-scaling the interval 60-305 days to unity, random effect-estimates of the intercept and DIM provide direct estimates of between-cow variability in peak and slope of lactation curves. Apparently a specification of parity-specific heterogeneity in the covariance structure eliminates the need to perform within-parity analyses. This test-day model consequently produces estimates of lactation curve characteristics for each individual cow that probably are useful for model validation at herd visits. However, this option has not yet been utilized at larger scale. Initial comparisons of fit different types of lactation curves (2-, 3-, 4-, or 5-parameter models) have indicated, however, that model-fit is improved considerably if the last milk test before drying off is excluded. The reason why this test-day often is an outlier is that the farmer often has started feeding the cow differently as part of the drying-off procedure. Perspectives: Multi-level across-herds models that will account for all the above mentioned effects like interaction between herd-level variables (e.g., treatment policy and culling strategy) and cow-level variables (e.g., SCC and milk yield potential) will become extremely complex and presentation of parameter estimates will not be straight forward. Currently, a meta-analytic approach that systematically combines results from multiple within-herd analyses seems more feasible.

[1]Enevoldsen C, Hindhede J, Kristensen T. Dairy herd management types. J. Dairy Sci. 1996; 79: 1221-36
[2]Nir-Markusfeld O, Enevoldsen C, Kroll O. Herd Health. Caesarea (Israel): Hachaklait, 1999: pages 7-8.
[3]Enevoldsen C. Indicadores de salud en el maejo de la produccion, in Proceedings III Jornadas Chilenas de Buiatria, Sociedad Chilena de Buitria, Osorno, Chile, 10 al 12 de Abril de 1997: pages 69-75.
[4]Internet-address: URL: http://www.epivetko.dk (in Danish)
[5]Kuhn MT, Freeman AE, Fernando RL. Approaches investigated to correct for preferential treatment. J. Dairy Science 1999; 82: 181