Epidémiol. santé anim., 1997, 31-32

EVALUATION OF CLINICAL MASTITIS AND SOMATIC CELL COUNT AS DIAGNOSTIC TESTS FOR SURVEILLANCE OF UDDER HEALTH IN DAIRY HERDS

Agger J.F.[1], Bartlett P.C.2, [1]Woudstra I, [1]Willeberg P, [2]Houe H, [1]Lawson L, [1]Enøe C

Les systèmes de surveillance des maladies dépendent de la performance des procédures de diagnostic utilisées. L'évaluation d'un test diagnostique nécessite toujours sa comparaison à un test de référence « gold standard », toutefois, souvent ces standards ne sont pas disponibles. A leur place, d'autres procédures épidémiologiques et statistiques peuvent être utilisées, telle la méthode du maximum de vraisemblance décrite par Hui et Walter (1980) et Walter et Irwig (1988). L'objectif de cette étude était d'évaluer les données sur les mammites et le taux cellulaire dans le lait comme test diagnostique pour la surveillance de la mamelle en élevage laitier. Les données étaient collectées en 1993-1994 sur 328 élevages, comprenant environ 17000 vaches. Une classification des animaux a été faite selon le taux cellulaire avec le seuil de 1 million (HSCC), les mammites cliniques (RMC), la parité (primipares vs multipares). La sensibilité, la spécificité, et la proportion d'animaux malades ont été estimées pour chacun des deux tests sans « valeur standard » en analysant les données à l'aide de l'option LE du logiciel BMDP. Certains élevages ont également été analysés avec une macro (em-algoritum) de SAS. Les résultats montrent quelques divergences entre les deux programmes. La sensibilité de HSCC (0.446) était plus élevée que pour RCM (0.312) et la spécificité avait la même amplitude dans les deux cas (0.926).
On conclut que la méthode du maximum de vraisemblance peut être utilisée pour estimer les paramètres du test diagnostique, et qu'il faudrait un choix judicieux des valeurs initiales. Les résultats ainsi que les comparaisons détaillées des méthodes d'analyse sont discutés.

All disease surveillance systems depend on the capabilities of the diagnostic procedures available. Evaluation of a diagnostic test usually requires comparison to a definitive test (gold standard), however, such standards are often not available. Examples of traditional test evaluation are *Agger et al.* (1985) evaluating the now former version of the Danish national mastitis control program at the herd level, and *Jørgensen et al.* (1982) evaluating the diagnostic values at different cut off levels of somatic cell counts in cow milk samples using bacteriology as the true state of nature. Instead, other epidemiological/statistical procedures can be used for test evaluation. For example, the maximum likelihood method has been described by *Hui & Walter* (1980) and *Walter & Irwig* (1988). Another method is the Bayesian approach which simultaneously infers the population prevalence, sensitivity, specificity, and positive and negative predictive values of one test at a time or 2 or more combined diagnostic procedures (*Joseph et al.* 1995). There are very few examples of estimating sensitivity and specificity of diagnostic tests by use of the maximum likelihood procedure, eg. *Spangler et al.* (1992). At this symposium, however, papers on the topic are also presented by *Willeberg et al.* (1997), *Chriel & Willeberg* (1997), and *Enøe et al.* (1997 a,b).

The decision to seek veterinary treatment of mastitis varies greatly among farmers. Some farmers consider even slight changes in the milk as requiring treatment for mastitis, while others do not call the veterinarian until severe disease is observed. Therefore, the sensitivity of recording clinical mastitis is very heterogeneous among farms. Automated somatic cell counting procedures are more independent of herd effects.

The purpose of this study was to evaluate data on clinical mastitis and cow somatic cell count as diagnostic tests for surveillance of udder health in dairy herds. Further, it was our purpose to study the relationship between sensitivity and specificity of these diagnostic tests in relation to herd size and herd average daily milk yield. The data, from the Danish Cattle Data Base, were collected during 1993-1994 from 2148 dairy herds. This preliminary study was restricted to 328 herds comprising a total of 154 101 cow tests for somatic cell count and two daily examinations at approximately 2 684 000 milkings in 17 000 cows for the occurrence of clinical mastitis.

MATERIALS AND METHODS

According to *Hui & Walters* (1980) the evaluation of 2 diagnostic methods using the maximum likelihood procedure requires data from 2 subpopulations with different proportions of disease occurrence. It is well known that the occurrence of mastitis increases with age, and data within each herd were therefore stratified into young cows (parity 1 and 2) and old cows (parity > 2). Within each of these groups, the data were cross-classified according to the results of the 2 diagnostic tests.

The first diagnostic test for mastitis was based on the record of a valid monthly milk test of a lactating cow with a SCC greater than 1 million cells per ml. The second diagnostic test for mastitis was defined on the basis of one or more veterinarian or farmer reports of clinical mastitis during the 1-year monitoring period from June 30, 1993 to July 1, 1994. Such reports could have originated from the veterinarian's financial and medical records system,

[1] Department of Animal Science and Animal Health, and
[2] Department of Clinical Studies, The Royal Veterinary and Agricultural University, Bülowsvej 13, DK-1870 Frederiksberg C, Denmark

or from the farm record system. In most instances, mastitis reports signified the administration of antibiotics which, in Denmark, require the attention of a veterinarian. The unit of analysis was the cow monthly milk test. Some cows contributed the maximum of 12 test-months to the database, and other cows which were culled or entered the milking herd after the beginning of the project contributed fewer than 12 test-months. The presence of a SCC value > 1 million cells per ml was considered a high SCC (HSCC) and was our first diagnostic test. Reported clinical mastitis (RCM) was the second diagnostic test, and was considered positive if a mastitis treatment had been recorded for that cow within 30 days before or after the date of the milk test. On this basis, each valid monthly milk test was declared to be positive or negative on the basis of each of the 2 diagnostic tests (RCM and HSCC). The combined RCM and HSCC test results for each test month for each cow (HSCC+ and RCM+, HSCC+ and RCM-, HSCC- and RCM+, or HSCC- and RCM-) were then summed for young cows (parity 1 and 2) and old cows (parity > 2) for each herd. Thus, a 2 X 2 table comparing RCM to HSCC was produced for young cows and a second 2 X 2 table was produced for old cows. These 8 numbers were the input into the maximum likelihood program (LE) in the software system BMDP (*Dixon WJ (ed)* 1995), and for 42 herds also into the SAS macro for the em-algoritm.

Six parameters were estimated: theta1 (the true proportion of mastitis among young cows), theta2 (the true proportion of mastitis among older cows), alpha1 (error rate among nondiseased by test1= HSCC), alpha2 (error rate among nondiseased by test2=clinical mastitis), beta1(error rate among diseased by test1), and beta2 (error rate among diseased by test2). The analyses for all 328 herds were carried out by 2 researchers using BMDP. As starting points for the iterations, one person consequently used initial values of 0.1, 0.1, 0.2, 0.2, 0.5 and 0.5 for the analysis of 170 herds, while the other researcher mainly used 0.05, 0.05, 0.05, 0.05, 0.85 and 0.85 according to the order of parameters mentioned above. If the program did not converge the first time for the individual herd, the results of the first iteration were entered into the program as new initial values in the search for the most likely set of parameter estimates. This process continued until a stable set of values was found, or until it was decided that a maximum could not be identified (convergence not possible).

It was assumed that samples from the same cow were independent, and that cows within the same herd were independent. It is also assumed that, conditional on the true disease state, the two tests on each individual are subject to independent errors, however, violation of these assumptions were not tested.

RESULTS AND DISCUSSION

Reliable results require reliable data and reliable analytical techniques. The first issue, therefore, was to critically consider the data and the methods. Hence, the reporting of these analyses will primarily address the data structure and the analytical methodology, and with less emphasis on the biological results.

Based on the initial values entered in BMDP(LE), the first researcher tended to receive high and stable estimates for theta1 and theta2. The first researcher never used zero or 1 as initial values, but rather substitute values of 0.01 and 0.999. The second researcher tended to get lower values for theta1 and theta2, and used zero as initial values but never 1 (only 0.999). This different approach tended to facilitate fast convergence. The high vs low results were probably not due to the different initial values chosen, but rather that the 2 persons had different preferences for initial values if the herd data did not converge. This showed that it was possible to manipulate the results, and therefore we are skeptical of the estimated parameters. Also, for several herds, it was possible to find more than one set of stable values for theta1, theta2, alpha1, alpha2, beta1 and beta2. Sometimes two sets of stable values were complementary to each other, e.g. (1 - alpha1) gave the estimate of alpha1. This is in agreement with *Hui & Walter's* (1980) results that the distributions often are symmetric.

Herds with zeros in one or more cells (empty cells) were approzimately 4 times as likely to not converge as were herds with no empty cells, when using the BMDP(LE) program. We therefore decided to exclude from further considerations in BMDP(LE) the 121 herds with one or more empty cells, ie. only the remaining 207 herds were evaluated in scatter plots and by estimating simple correlation coefficients. Obviously, such exclusion inflated the overall estimates of theta1 and theta2.

A scatter plot of theta1 against theta2 showed that theta1 estimates were divided in 2 clusters. The clustering, however, was less pronounced for theta2. The majority of herds had small estimates of both parameters and a cluster of 42 herds had theta1 and/or theta2 greater than 0.6. The 42 herds were analysed using the SAS macro em-algoritm. Twentyeight of 33 herds (85%) with theta1 above 0.6 had new estimates below 0.6, and 23 of 26 herds (88.5%) with theta2 above 0.6 had new estimates below 0.6. The new estimates were often complementary to the results obtained in BMDP. This clearly indicates the need to analyse all the herds using the SAS macro em-algoritm. The estimates of sensitivities obtained in SAS had a greater dispersion for HSCC and especially for RCM as compared with estimates obtained in BMDP. This greater dispersion seems more biologically plausible as compared to the more clustered BMDP-results. The specificities of HSCC in the 42 herds changed from a wide dispersion in BMDP to a cluster above 0.6 in SAS. Specificities for RCM also changed to higher, more biologically plausible levels in SAS as compared with BMDP. The analysis of 2 herds using the gibbs sampling technique resulted in several maxima, confirming the result from BMDP that several maxima can be identified by continuing to enter new initial values. Using the em-algoritm in SAS, the complementary values to the BMDP results were estimated, and often alphas in BMDP were equal to 1-beta in SAS. It can be concluded that, in our analysis, no specific biological conclusions can be reached using only the BMDP(LE) program. However, in relation to herdsize and milk yield, the scatter plots suggested that neither herd size nor herd average daily milk yield was clearly related to any of the parameters (se1, se2, sp1, sp2, theta1 and theta2). Regression analyses, however, showed that herd average daily milk yield was nearly significantly related (p=0.065) to the specificity of clinical mastitis, herd size is almost significantly related (p=0.065) to the sensitivity of HSCC, and significantly related (p=0.04) to the sensitivity of RCM. If this is correct, then it may not be true that farmers of high yielding herds are better in recording diseases than small and lower yielding herds. This may be

new information in the old discussion regarding the relationship between milk yield and occurrence of disease. The mean estimates of the 6 parameters for 207 herds analysed in the BMDP(LE) program are presented in Table I. The sensitivities of HSCC is, as expected, higher than for RCM, and the specificities of the 2 tests are similar. As expected we also found a lover proportion of mastitis in young than old cows. The level og the proportion can, though, be questioned. The average milk yield and hard size are given also.
Some of the herds had a low sample size. However, herds that did not converge were in the range of 300 - 700 herd tests, which was not different from the distribution among converged herds. The use of 1 mill SCC per ml may not be the best cut-off value. Also, the interval of 30 days on each side of the HSCC may be to long; it should be noted that intervals overlap for consequtive months.

Table I
Mean values of 207* herds for HSCC and RCM as diagnostic tests for mastitis, and mean herd size, and average daily milk yield

Variable	Mean	Std Dev
HSCC sensitivity (1- beta1)	0.446	0.366
RCM sensitivity (1 - beta2)	0.312	0.294
HSCC specificity (1 - alpha1)	0.927	0.160
RCM specificity (1 - alpha2)	0.926	0.125
Proportion in young cows (theta1)	0.199	0.318
Proportion in old cows (theta2)	0.268	0.270
Average daily milk yield (ecm/d)	23.02	2.164
Average herd size (milking cows)	82.9	38.2

*Herds with empty cells in the HSCC and RCM comparison have been excluded, as have herds which failed to converge.

REFERENCES

Agger J.F., Willeberg P., Schmidt Madsen P., Klastrup O., 1985. A case control study of environmental and managerial factors in the Danish national mastitis control program. Proceedings 4th ISVEE, Singapore, November, 438-440.

Chriél M., Willeberg P., 1997. Dependency between sensitivity, specificity and prevalence analysed by means of gibbs sampling. Proceedings 8th ISVEE, Paris, France.

Dixon W.J. (ed), 1995. Maximum likelihood estimation (LE). BMDP Statistical Software Manual version 7.0, University of California Press, 775-793.

Enøe C., Andersen S., Thomsen LK., Mousing J., Leontides L., Willeberg P., 1997a. Estimation of the sensitivity and specificity of two diagnostic tests for the detection of antibodies against Actinobacillus pleuropneumoniae type 2 in pigs by maximum-likelihood estimation and gibbs sampling. Proceedings 8th ISVEE, Paris, France.

Enøe C., Mousing J., Willeberg P., 1997b. Chronic pleuritis in pigs for slaughter: Analysis of infectious and rearing system related risk factors, adjusting for extrabinomial variation in the data. Proceedings 8th ISVEE, Paris, France.

Hui S.L., Walter S.D., 1980. Estimating the error rates of diagnostic tests. Biometrics 36, 167-171.

Joseph L., Gyorkos T.W., Coupal L., 1995. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Amer. J. Epidemiol, 141, 263-272.

Jørgensen J.C., Schaap P., Agger J.F., Willeberg P., Schmidt Madsen P., 1982. Celletallet i koprøver anvendt i mastitisdiagnostik og -bekæmpelse (Somatic cell count in cow milk samples used in matitis diagnostics and control). Proceedings 14th Nordic Veterinary Congress, Copenhagen, July 1982, 268-269.

Spangler E., Bech-Nielsen S., Lawrence E.H., 1992. Diagnostic performance of two serologic tests and fecal culture for subclinical paratuberculosis, and associations with production. Prev. Vet. Med. 13, 185-195.

Walter S.D., Irwig L.M., 1988. Estimation of test error rates, disease prevalence and relative risk from misclassified data: A review. J. Clin. Epidemiol 41, 923-937.

Willeberg P., Wedam J.M., Gardner I.A., Holmes J.C., Mousing J., Kyrval J., Enøe C., Andersen S., Leontides L., 1997. A comparative study of visual and traditional post mortem inspection of slaughter pigs: Estimation of sensitivity, specificity and differences in non-detection rates. Proceedings 8th ISVEE, Paris, France.