J. Dairy Sci. 90:5415–5427
doi:10.3168/jds.2007-0148

Estimating Degree of Mastitis from Time-Series Measurements in Milk: A Test of a Model Based on Lactate Dehydrogenase Measurements

N. C. Friggens,*[1] M. G. G. Chagunda,*[2] M. Bjerring,* C. Ridder,*† S. Højsgaard,* and T. Larsen*
*University of Aarhus, Faculty of Agricultural Sciences, Research Centre Foulum, PO Box 50, 8830 Tjele, Denmark
†Lattec I/S, Slangerupgade 69, 3400 Hillerød, Denmark

ABSTRACT

The aim of this study was to test a model for mastitis detection using a logic that allows examination of time-related changes and a progressive scale of mastitis state (i.e., not using specificity/sensitivity). The model produces a mastitis risk (MR) for individual cows on a scale from 0 (completely healthy) to 1 (full-blown mastitis). The main model input was lactate dehydrogenase (LDH; μmol/min per L) × milk yield. Test data containing 253 mastitis cases were used. Proportional samples were collected from each cow at each milking and analyzed for LDH and somatic cell count (SCC). The basis for the health definitions was veterinary treatment records. A refinement of the basic health definitions was made using systematic positive deviations in log(SCC) to indicate untreated infections. Two subsets of cows were identified: mastitic cows and cows completely free of mastitis (healthy controls). The time-profiles of these 2 groups in a 60-d window relative to day of veterinary treatment were examined. Model reliability throughout all stages of lactation and degrees of infection was examined using SCC as a continuous measure of degree of mastitis. The time-profile for the health controls was flat throughout the 60-d window with a median MR of 0.02. In contrast, the profile of the mastitic cows increased above the control cows' baseline from about –6 d, rising to a MR value of 0.20 at d 0, and declining to the control level after treatment. There were significant differences between mastitic and healthy cows from –4 to +2 d relative to veterinary treatment. When cases were time-aligned to peak of infection, rather than veterinary treatment, there was a much sharper peak to the time-profile of mastitic cows. The median MR at peak was 0.62 and the mean was 0.80. Using these data, the MR value of 0.62 had a <1% likelihood of actually coming from a healthy control. Testing against SCC, on the whole data set, showed that only 2.1% of all MR values had an error >0.7. These estimates of model reliability are comparable with the greatest values reported in the literature and, additionally, the model was able to detect significant differences between mastitic and healthy cows 4 d before treatment. It was also found that specificity/sensitivity calculations are inappropriate for evaluating time-related changes and a progressive scale of predicted mastitis state.

Key words: mastitis, detection, time-series, dairy cow

INTRODUCTION

Systems for detecting mastitis have become increasingly sophisticated (Cavero et al., 2007) reflecting advances in on-farm technology as well as an increasing focus on milk quality and animal welfare. This increasing sophistication raises a number of issues with respect to testing such models. The traditional way of evaluating the efficacy of a disease detection system or a treatment is by using specificity/sensitivity calculations. Reliable estimates of the diagnostic sensitivity and specificity are needed in many settings (Greiner and Gardner, 2000). In the field, veterinary practitioners require sensitivity and specificity estimates to update clinical inferences, and considerations of accuracy are important in test selection. Surveillance programs need sensitivity and specificity estimates for sample size calculations, and epidemiologists and risk analysts use them to adjust prevalence estimates for misclassification and to parameterize decision-trees (Greiner and Gardner, 2000). Specificity and sensitivity calculations, however, have major limitations in the context of some of the emerging features of newer mastitis detection systems.

The system on which this study focuses is designed to use repeated measures of an indicator of mastitis; that is, rather than detecting mastitis from one isolated measurement, it claims to be able to track the develop-

Received February 27, 2007.
Accepted August 17, 2007.
[1]Corresponding author: n.friggens@agrsci.dk
[2]Current address: Scottish Agricultural College, Sustainable Livestock Systems Group, Dairy Research Centre, Midpark House, Bankend Road, Dumfries, DG1 4SZ, Scotland.

ment of mastitis through time. Implicit in this approach is the concept of a degree of mastitis, or risk of mastitis, that can evolve through time. Concepts have an intrinsic scientific value but when they are incorporated into a model that is designed for application, it is important to evaluate their worth in terms of the practical advantages they may provide the end-user. What are the advantages for mastitis detection of tracking the development of a continuous measure of mastitis through time? This question is difficult to address with specificity/sensitivity testing, in which it is assumed that mastitis state is a binary variable (healthy or sick), and the time dimension is reduced to an aggregate value within a fixed time-window. It is possible to overcome these constraints, at least in part, by performing a series of specificity/sensitivity calculations that track through a sequence of time-points or work through a series of different thresholds for classifying cows as sick or healthy; for example, using receiver operating characteristic methodology (Lasko et al., 2005; Gardner and Greiner, 2006). This approach, however, largely ignores the correlations inherent in the progression of a time-series of measurements. Further, the limited uptake of these methodologies indicates that they tend to produce results that are intuitively difficult for nonspecialists to interpret. Thus, the aim of this study was to test a dynamic model for mastitis detection using a rationale that allows explicit examination of the time-related changes and a progressive scale of predicted mastitis state.

MATERIALS AND METHODS

The Model Being Tested

The model being tested in the current analysis has been described in detail by Chagunda et al. (2006). This model produces an overall risk of acute mastitis for individual cows, on a scale where 0 = completely healthy and 1 = full-blown clinical mastitis. It is a dynamic and deterministic hybrid system and is designed to run in real-time; that is, each time a new input occurs. The main model input is lactate dehydrogenase (**LDH**) amount (LDH activity × milk yield). After smoothing and filtering using an extended Kalman filter, this is used to generate an indicator-based risk (**IBR**). In the IBR, the increase in LDH amount is calculated relative to a baseline for that cow. This baseline is a rolling average of the previous 7 d of values of LDH and allows the model to distinguish between acute increases in LDH amount and generally high levels associated with chronic mastitis. Other animal- and herd-related factors available on-farm (DIM, udder characteristics, milking duration, and disease history) are used to generate an additional risk factor (**ARF**). The overall risk of acute mastitis is a weighted sum of the IBR and the ARF in the ratio 1:0.25. The model tested in the present study contained the following minor modification relative to the model described in Chagunda et al. (2006). The function calculating the component of mastitis risk resulting from the slope in LDH amount was originally linear (the greater the slope, the greater the mastitis risk) but is now a sigmoid function. This prevents the extreme values that occur if slopes are calculated over very short time intervals; for example, if a milking is interrupted. Over the normal range of slope values, the modification has a negligible effect.

The model has many features that warrant testing, especially in relation to the ARF. In the present study we have chosen to focus on the overall risk of acute mastitis, because this is the major output and the one by which the usefulness of the model in commercial application will be primarily judged. For the purposes of testing, no constraints were placed on the values of overall risk of acute mastitis. Although the model was designed to produce values in the range from 0 and 1, values can be outside this range because, for example, a negative slope of LDH amount generates a negative component to the IBR.

Structure of Test Data

A test data set was created by a period of intense monitoring and milk sampling on one farm, the Danish Cattle Research Centre (Tjele, Denmark). All cows were milked with an automatic milking system (3 units, on average 2.3 ± 0.83 milkings/cow per d) in which the milk yield was automatically recorded. The data used were collected in the period between April 2003 and September 2006. The cows were of 3 breeds: Danish Holstein (n = 123), Danish Red (n = 130), and Jersey (n = 79), and were loose-housed in one barn. After data editing (see below), the total number of lactations in first, second, and greater parities were 246, 179, and 177, respectively. Proportional samples of composite milk (all quarters combined) were collected automatically from each cow, at each milking. The automatic milk sampling system was emptied in the morning and in the afternoon. Milk samples were kept at 4 C until laboratory analysis, which was done within 24 h of sampling.

Milk samples were analyzed for LDH and SCC. Samples were analyzed for LDH activity (μmol/min per L) using a fluorometer (Fluostar, BMG Labtechnologies, Offenburg, Germany) as described by Larsen (2005). Somatic cell count was measured using Fossomatic 5000 automatic equipment (Foss Electric, Hillerød, Denmark). Additionally, foremilk quarter samples were taken each time a veterinary treatment was initiated

and cultured, and the resulting proportions of the different pathogen types were typical of Danish dairy farms (36% *Staphylococcus aureus*, 25% CNS, 10% *Streptococcus uberis*, 5% *Escherichia coli*, 5% *Streptococcus dysgalactiae*, and 13% no culture).

Test Rationale

The rationale for testing was according to the following stepwise progression, each step being conditional on the outcome of the previous step.

First, 2 subsets of cows were identified. These were cows that could be classified, with reasonable certainty, as either having mastitis or being completely free of mastitis (herein called healthy controls) within specific time-windows. The health definitions involved and the procedure for selecting control cows are described below. The profiles of overall risk of acute mastitis (subsequently referred to as mastitis risk, **MR**) within the time-window of these 2 groups were then examined to see if they were significantly different. The test hypothesis was that there would be a significant difference between mastitic and healthy controls at and around the time of veterinary treatment, and that this difference would decrease with increasing time (before and after) from treatment.

Second, given a clear description of the time profile of a typical acute mastitis case, the ability of the model to provide early identification of mastitis cases was explored. This was done in terms of both the time lag between model identification and veterinary treatment and the degree of uncertainty associated with different levels of model-generated MR. The latter provides a measure of the reliability of the model based on the time windows.

Third, the reliability of the model throughout all stages of lactation and degrees of infection was examined. To do this, a continuous reference measure of degree of mastitis is required, and it was assumed here that SCC provides such a measure (the pros and cons of this assumption are considered in the Discussion). Thus, the reliability of the model was quantified relative to SCC.

Health Definitions

The basis for the health definitions was veterinary treatment records. In Denmark, only veterinarians are permitted to initiate treatments and they are legally required to record treatment data. Cows were selected for veterinary examination by the farm staff based on cows showing physical signs of mastitis and from weekly visual inspection of milk. A mastitic cow was one that had a recorded veterinary treatment of mastitis with the day of treatment being designated days from mastitis = 0. From the start of lactation until 30 d before the first recorded treatment of mastitis, cows were classified as healthy and were available for selection as control cows.

Because of the possibility of mastitis cases being overlooked or misdiagnosed, a refinement of the basic health definitions was made. This was based on patterns of SCC through time. When a positive systematic deviation in a cow's time-series of log(SCC) greater than 2.25 occurred the cow was assumed to have had a mastitis and to not be healthy in the period –10 to +10 d relative to the SCC peak. (The method for calculating the deviations is described below). The threshold of 2.25 was chosen because the average increase in logSCC for cows identified as having had mastitis on the basis of this threshold was the same as the average increase in log(SCC) for cows identified by the veterinarian as having had mastitis. If one considers the veterinary treatment records relative to the SCC peaks in terms of specificity and sensitivity, the basic health definitions can be subdivided within the classes mastitic and healthy controls as follows. If there was a veterinary treatment record and an SCC peak (true positive), then the case was termed "true mastitic"; if there was a veterinary treatment record but no SCC peak (false positive), then the case was termed "false mastitic"; if a healthy control had no SCC peak during the control period or previously in that lactation, it was termed "true healthy"; and finally, if a healthy control had an SCC peak during the control period or previously in that lactation, it was termed "false healthy". The choice of an SCC peak of 2.25 resulted in 76% of the cases treated by the veterinarian being classified as true mastitic and 35% of the healthy controls being classified as true healthy.

Data Editing

The raw data consisted of 496,014 milking records, which included 101,046 records with LDH measurements. There were 749 cow lactations with at least 1 LDH measurement. Data were edited for abnormal frequency of sampling. If the gap between 2 consecutive LDH amount records was greater than 5 d, then the milk records in that period were coded as being in a gap. The same applied between calving and the first LDH record, and between the last LDH record and the last record in a lactation. Further, limited runs of valid LDH data (i.e., <10 d between 2 gaps) were deemed to be of little use and also coded as being in a gap; in total, 6,922 LDH records occurred in gaps. Using the gap code as a basis, lactations with few, spread out LDH measurements were excluded. This was done by calcu-

lating for each run of nongap data, the frequency of LDH measurements and the number of LDH measurements so that the exclusion was based on the nongap windows. If (and in this order) the total number of nongap LDH records for the whole lactation was less than 21, then that lactation was excluded. If, additionally, the maximum frequency of LDH measures in any nongap (for a given cow-lactation) was less than 0.35 LDH measures per milking, then that lactation was excluded. If, additionally, the maximum number of LDH measures in any nongap was less than 21, then that lactation was excluded. The choice of gap lengths was based on consideration of the typical time-course of a mastitis case combined with visual inspection of the data. The resulting data set contained 602 cow lactations, 90,477 LDH measurements, and 253 mastitis treatment records. In the selection of mastitis cases, if there was a gap in LDH during the period from –30 to +30 d from the day of veterinary treatment, then that case was excluded.

If 2 records of veterinary treatment of mastitis occurred within 8 d, they were treated as a single case (IDF, 1997). For each mastitis case, days from treatment was calculated. Days from treatment was positive and increased with time after the mastitis case, whereas it was negative and more negative with time before the mastitis case. Between 2 mastitis cases, the change from positive days from treatment (after the previous case) to negative days from treatment (before the next case) occurs midway between the 2 cases. In this study, a mastitis case window was defined as being from –30 to +30 d from treatment. For each mastitis case, matching control periods were chosen from the pool of cows of the same breed and lactation number class (1, 2, 3+) that had no veterinary treatment of mastitis in that lactation until at least 30 d after the DIM of the mastitis case being matched. The chosen control period spanned the same DIM of the case being matched. Mastitis cases occurring after 350 DIM were excluded. The process of finding healthy controls was carried out in 3 rounds, a round consisted of finding 1 healthy control for each mastitis case, where possible. The second round repeated the process to identify a second healthy control for each mastitis case, where possible, and likewise for the third round. Thus, a maximum of 3 healthy controls were found per mastitis case. A given cow lactation was allowed to provide more than 1 control period (each to different mastitis cases) but if a cow lactation was used in 1 round of selection, then it was excluded from subsequent rounds of selection. Within cow lactation, control periods were not permitted to overlap. The final number of mastitis cases was 185 (from 146 cow lactations). The final number of healthy controls was 426 (from 319 cow lactations). When comparing true mastitic vs. true healthy cows, to maintain balance in the distribution of cases relative to DIM, breed, and parity, it was decided to exclude whole sets of mastitic cows and their controls when one or other was not classified as "true" (mastitic or healthy). For the "true" comparisons, this resulted in a data set containing 58 true mastitic cases and 71 true healthy cases.

The following procedure was used to identify peaks in the time-series of SCC. For each time-point t, the sum of all the individual changes in log SCC between t-25 milkings and t was calculated. This rolling cumulative sum is relatively insensitive to random noise but increases when there is a consistent increase in log SCC over a number of measurements. Each individual change in log SCC was calculated as

$$(\text{logSCC}_t - \text{logSCC}_{t\text{-}1}) - (\text{expected change in logSCC}_{(t\,-\,t\text{-}1)} \text{ for healthy cows of that breed-parity}).$$

The breed-parity expected change; that is, the normal lactation curve in log SCC, was calculated by fitting an exponential + linear effect to the average log SCC vs. DIM for healthy cows as follows. For each breed-parity combination independently, the daily average log SCC was regressed on DIM using data from 50 to 350 d to obtain the linear trend through lactation and adjust for it. For DIM 0 to 50, the rapid decline in log SCC was modeled as an exponential decline estimated by regressing $\log_e$(adjusted average daily log SCC values) on DIM. The resulting exponential + linear equations provided the breed-parity expected change in $\text{logSCC}_{(t\,-\,t\text{-}1)}$. In practice, these expected changes were very small relative to the changes in logSCC related to known mastitis cases. The same rolling cumulative sum procedure was used to identify peaks in the time-series of MR (although in this case the expected change was assumed to be zero). These were then subsequently used when time-aligning the mastitis cases according to the time of peak mastitis risk in the window –10 to +5 d from treatment.

Statistical Analysis

The time-points at which the MR profile of the true mastitic cows was significantly different from that of the true healthy cows were analyzed using a mixed model (SAS version 8.e; SAS Inst. Inc., Cary, NC) with breed, parity, days from treatment, and the interaction between days from treatment and case-type (healthy vs. mastitic) as fixed effects. The model used was

$$\text{MastRisk}_{ijklm} = a + \text{Breed}_i + \text{Lakclas}_j + \text{Day}_l + \text{Case} \times \text{Day}_{kl} + A_{ijkmn} + \varepsilon_{ijklm}$$

where MastRisk is the MR of cow m (m = 1...110) of breed i (i = Danish Red, Danish Holstein, Jersey), lactation class j (j = 1, 2, 3+), case-type k (k = healthy, mastitic) on day from treatment l (l = –30,...,30); a is the intercept, A_{ijkmn} is the random cow effect, and ε_{ijklm} is the residual. It was assumed that $A_{ijkmn} \approx N(0, \sigma^2_A)$, $\varepsilon_{ijklm} \approx N(0, \sigma^2_\varepsilon)$, and all are independent. To account for repeated measures and unequal variances in the healthy and mastitic profiles at different time-points, cow within parity was included as a random effect within combinations of case-type k and period n with 6 levels defined as the intervals [–30, –20], [–20, –10], [–10, 0], [0, 10], [10, 20], [20, 30] days from treatment. The reduction of days from treatment into these 6 periods in the random effect was necessary to obtain convergence of the estimation procedure. All effects were included as factors (i.e., not as continuous variables), and Satterthwaite's method for estimating degrees of freedom was used. Not including case-type as a main effect allowed direct estimation, for each day, of the difference between healthy and mastitic cows and the significance of this difference.

The reliability of the model for distinguishing true mastitic from true healthy cows within the time window was estimated using the mean (μ) and the standard deviation (σ) of MR from the true healthy cows to calculate the likelihood of a given MR (x) belonging to the normal standard distribution of the true healthy cows, thus:

$$\text{likelihood} = 1 - F[(x - \mu)/\sigma]$$

where F is the distribution function for the standard normal distribution.

The mixed model procedure was also used to derive the residual MR values from the whole data set (not just the windows around mastitis treatments) after adjustment for the corresponding SCC values. Because there is no a priori linear relationship between MR and log SCC, the second and third powers of log SCC were included in the regression, as were milk yield and log SCC × milk yield to allow for the fact that MR is based on LDH amount (LDH concentration × milk yield), whereas log SCC is not:

$$\text{MastRisk}_{ij} = a + b \times \text{logSCC}_{ij} + c \times (\text{logSCC}^2_{ij}) + d \times (\text{logSCC}^3_{ij}) + e \times \text{MY}_{ij} + f \times (\text{logSCC}_{ij} \times \text{MY}_{ij}) + \varepsilon_{ij}$$

where MastRisk is the MR of cow i on day of lactation j; logSCC is the log of SCC; MY is milk yield at that milking (kg); a is the intercept; b,c,d,e,f are the regression coefficients; and ε_{ij} is the residual. In this regression, logSCC, $(\text{logSCC})^2$, $(\text{logSCC})^3$, milk yield, and the interaction between milk yield and logSCC were included as continuous explanatory variables for MR. Cow within parity was included as a random effect.

RESULTS

Figure 1 shows a typical example of a time profile for 1 cow through lactation of the LDH amount and the model-derived MR. In this example, the cow received veterinary treatment for mastitis on d 73 and 90 after calving. The corresponding profile of SCC and milk yield is also shown. To make valid comparisons between healthy control cows and mastitic cows, these time profiles were aligned based on days from veterinary treatment.

Time Profiles of Mastitic and Healthy Cows Relative to Days from Veterinary Treatment

The time profiles of the median MR of mastitic and healthy control cows are shown in Figure 2A. These are based solely on the presence or absence of veterinary treatments. The profile for the health controls is flat throughout the 60-d window with an average median MR of 0.02. In contrast, the profile of the mastitic cows increased above the control cows' baseline from about –8 d from treatment, increasing to a MR value of 0.12 at –4 d from treatment and remaining at that level until d 0. After treatment, the MR declined rapidly to a level below that of the healthy controls, from which it gradually increased to the control cow level. The corresponding 25 and 75% quartiles of the distribution of MR are shown in Figure 2B.

It is clear from Figure 2 that there were systematic differences between the time profiles of median MR of mastitic and healthy cows and that these differences were related to days from treatment. Thus, the model can detect clinical mastitis. The peak of the time profile of median MR for mastitic cows was, however, only 0.12 and at no time was there separation between the 25 and 75% quartile ranges of the mastitic and healthy cows. A possible reason for this could be that the correspondence between veterinary treatments and true mastitis is not 100%. Therefore, as described in Materials and Methods, peaks in SCC were used to differentiate veterinary-treated cows into true mastitic and false mastitic (141 and 44 cases, respectively). Likewise, control cows were differentiated into true healthy and false healthy (147 and 279 cases, respectively). The median time profiles of true healthy and true mastitic cows are shown in Figure 3. The overall patterns of these time profiles are similar to those based solely on veterinary treatments but the peak for true mastitic cows is better defined, increasing to a higher peak of 0.2 on d 0. There

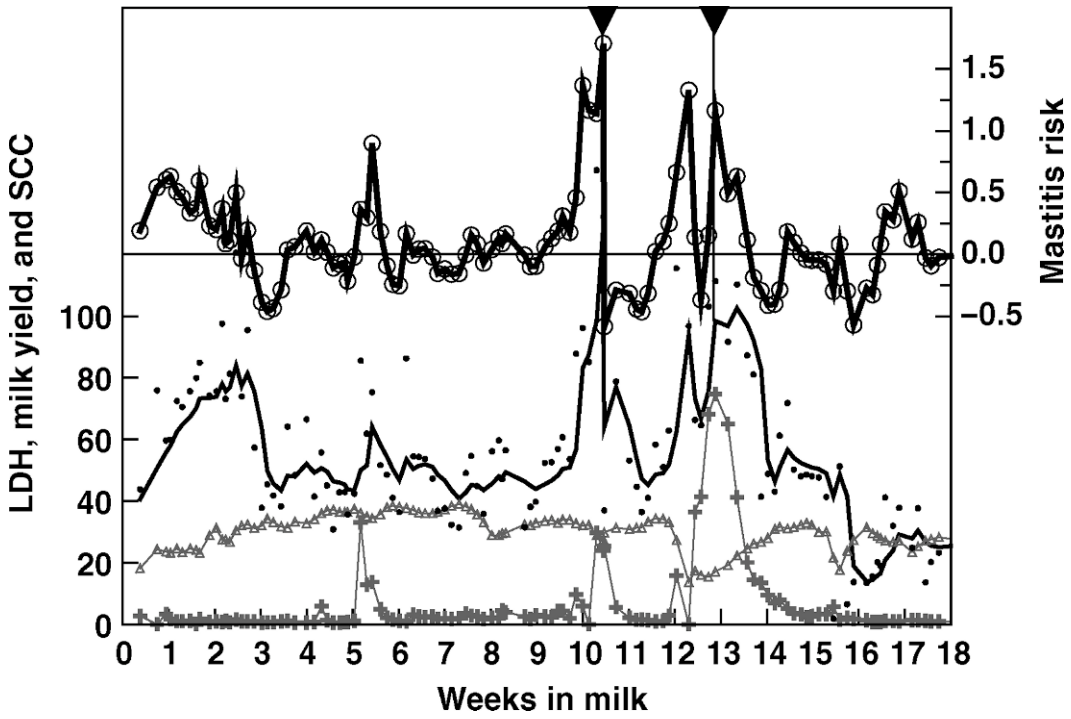

Figure 1. An example of an individual cow mastitis risk (MR) profile (○; right y-axis) relative to weeks in milk. Peaks in MR (e.g., at wk 5.5, 10.5, 12.3, and 12.8) are generated by rapid increases in the model indicator, smoothed lactate dehydrogenase (LDH) amount (μmol/min; left y-axis), as shown by the solid line with no symbols. The raw, unsmoothed, values of the inline indicator LDH amount are shown as solid circles. This cow received veterinary treatment for mastitis on DIM 73 and 90 (indicated by arrow, ▼) but not on DIM 39 (wk 5.5) when both MR and SCC showed a peak. The profile of SCC is shown on the left y-axis (+, count × 10^{-5}); milk yield (△, kg/d) is also shown on the left y-axis.

were significant differences between true mastitic and true healthy cows from –4 to +2 d from veterinary treatment (Figure 4). The average MR for true healthy cows across the time window was 0.044, which was not (quite) significantly different from 0 ($P = 0.056$). There were no days on which the average MR of the true healthy cows differed significantly from the average of 0.044. The MR of Jersey cows was, on average, 0.028 less than the other breeds ($P < 0.05$). There was no significant effect of parity on MR. Despite a significant difference between mastitic and healthy cows from –4 to +2 d from veterinary treatment, the peak for mastitic cows was low (Figure 4) when considering that the intended scale for MR was from 0 for healthy cows increasing to 1 for clinical mastitis.

Time Profiles of Mastitic and Healthy Cows Aligned to Peak in Model-Derived MR

To this point, the time profiles were aligned according to the day of veterinary treatment. The implicit assumption in this method of alignment is that there was a constant lag between time of occurrence of mastitis and time of veterinary treatment. If this were not the case, then it would be expected that the MR peaks for different cows would be horizontally offset from one another (i.e., on the time from treatment axis). It then follows that when the median time profile is calculated, its peak will be reduced because at that time, a proportion of the cows would be in postpeak decline (and a proportion would be in prepeak incline). Because there was no a priori reason for assuming a constant lag between peak and treatment, the time profiles of the mastitic cows were realigned according to the peak of each profile (within ±5 d of veterinary treatment, see Materials and Methods). On average, the peak MR occurred 1.0 d before veterinary treatment, with a large standard deviation of 2.8 d.

The peak-aligned time profiles for true mastitic and true healthy cows are shown in Figure 5. As expected, the effect of aligning to peak MR was to produce a much sharper peak to the time profile of true mastitic cows. The median MR at peak was 0.62 and the average was 0.80. The time profile of MR was largely symmetrical around the time of peak although after peak it tended to drop below the level of the healthy controls (Figures 4 and 5). This reflects the way in which the MR was calculated from the indicator, LDH amount. The IBR

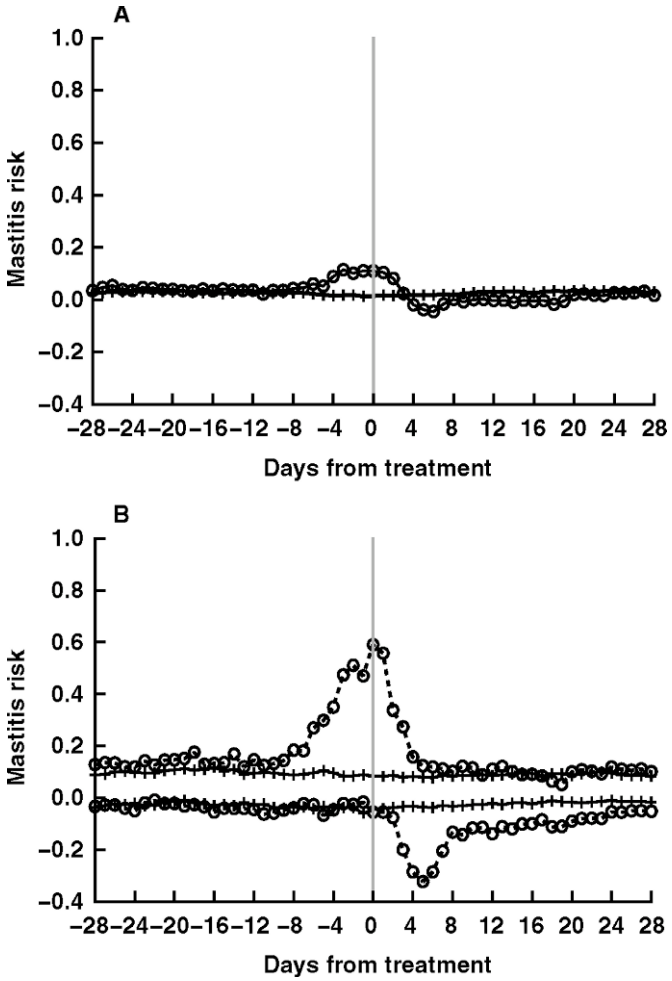

Figure 2. Mastitis risk (MR) relative to days from veterinary treatment of mastitic cows (○) and healthy controls (+), where health definitions are based solely on the presence or absence of veterinary treatments. Panel A shows median values and panel B shows the 75 and 25% quartiles of the distribution of MR values. Mastitis risk was generated by a model based on changes in the level and slope of milk lactate dehydrogenase amount.

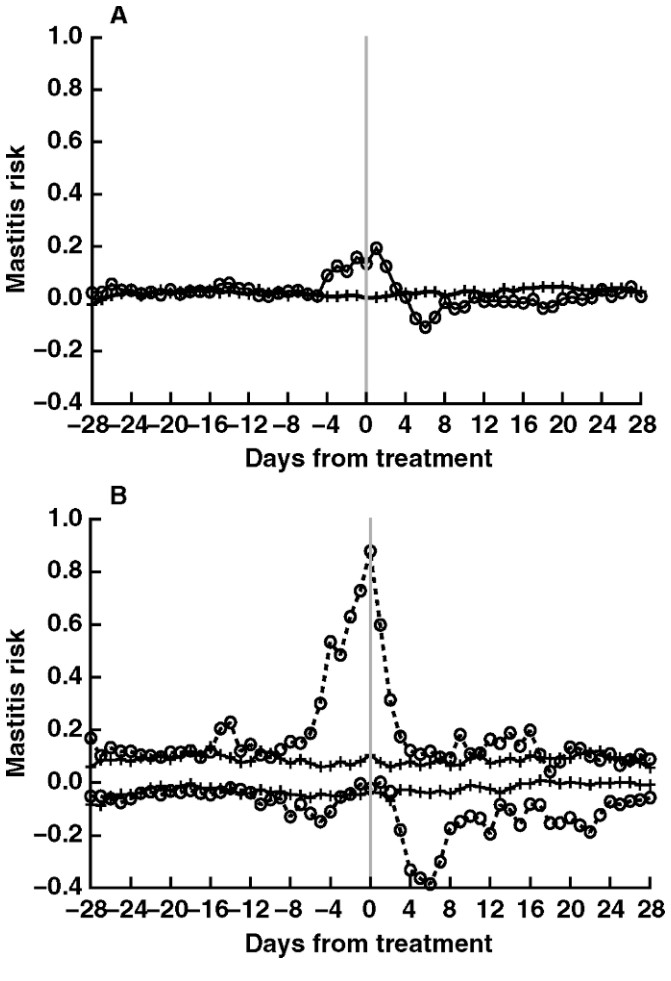

Figure 3. Mastitis risk (MR) relative to days from veterinary treatment of true mastitic (○) and true healthy controls (+), where health definitions are based on the presence or absence of veterinary treatments supplemented by changes in SCC (true healthy controls had no veterinary treatment and no peak in SCC; true mastitic cases had a veterinary treatment and a peak in SCC). Panel A shows median values and panel B shows the 75 and 25% quartiles of the distribution of MR values. Mastitis risk was generated by a model based on changes in the level and slope of milk lactate dehydrogenase amount.

has 2 additive components, one based on the slope and one based on the level of the LDH time-series [see Chagunda et al. (2006) for details]. The steeper the slope in LDH, the greater the slope-based risk. When the slope is negative, so is the slope-based risk; this allows the model to differentiate between 2 cases with the same level of LDH but where one is prepeak and rising and the other is postpeak and falling. Because the MR is designed to capture acute mastitis cases, the level-based risk uses not the absolute LDH level but the current level minus a baseline for that cow. This baseline is a rolling average of the previous 7 d of values of LDH. Thus, once the LDH level has returned to baseline postmastitis, the level-based risk is zero, and the IBR can be negative if there is still a downward slope in LDH. This is why the median MR for mastitic cows goes below that of the healthy controls postpeak. In practice, this may not be sensible, especially from the point of view of presentation to the end-user, and setting bounds (0 to 1) for MR would be one way to deal with this. In the present study, however, it is desirable to not truncate the MR values when evaluating the underlying equations. Had the MR values been truncated, then the estimates of the variability associated with the median profile would have been biased, which would have affected the quantification of the reliability of using any given value of MR to indicate mastitis.

For the peak-aligned time profiles of MR, the variation within case-type in MR values on any given day

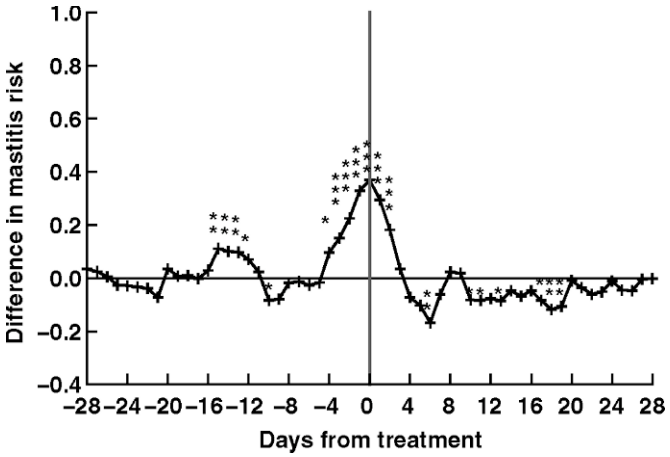

Figure 4. The average difference in mastitis risk profile between true mastitic and true healthy cows. Days on which the difference was significant are indicated by 1, 2, or 3 vertically aligned asterisks (*) for $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively. Mastitis risk was generated by a model based on changes in the level and slope of milk lactate dehydrogenase amount. True healthy cows had no veterinary treatment and no peak in SCC; true mastitic cows had a veterinary treatment and a peak in SCC.

was close to being normally distributed. This permitted calculation of the likelihood of a MR value being part of the normal distribution of true healthy cows. The average median value for the healthy cows across the time window was 0.024, and the average standard deviation was 0.221. Based on these estimates of the healthy population, the dashed line in Figure 5 gives the likelihood of a given risk value being part of the healthy population. For example, a MR value of 0.6 has a <1% likelihood of actually coming from a healthy control, and a MR value of 0.3 has a 15% likelihood of coming from a healthy control (Figure 5). Likewise, when the MR value was close to 0 and similar to that of the control cows, the likelihood of this value belonging to the healthy population was 50%; that is, there is an equal chance of a MR value of 0 belonging to healthy cows or to cows that will subsequently have mastitis. This likelihood provides a measure of the uncertainty of diagnosis should one use a particular MR value as a threshold for taking action; for example, calling the veterinarian or starting a treatment.

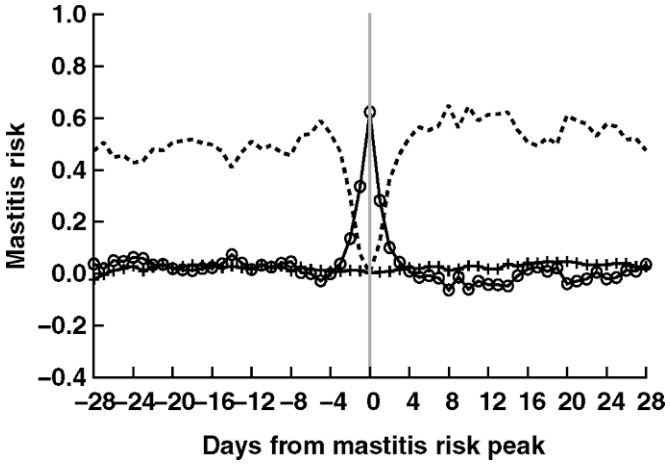

Figure 5. The median values of mastitis risk (MR) time-aligned according to peak in mastitis risk rather than time of veterinary treatment (mastitic cows: —○—; healthy cows: —+—).The likelihood of the median MR value of the mastitic cows belonging to the healthy population is shown by the dashed line (on the same scale as MR). Mastitis risk was generated by a model based on changes in the level and slope of milk lactate dehydrogenase amount. True healthy cows had no veterinary treatment and no peak in SCC; true mastitic cows had a veterinary treatment and a peak in SCC.

Figure 6 shows the MR peak-aligned time profiles for SCC, milk yield, and the 2 components of the MR: the IBR and the ARF. Figure 6 shows clearly that the MR profiles for true mastitic cows corresponded closely to the expected changes in SCC and milk yield. Further, it can be seen that the basis of the MR profile comes from the IBR, which is the rate of change and level of LDH amount. The ARF shows no clear time trend relative to mastitis peak.

Model Performance Relative to SCC Measurements

The shape of the time profiles of MR and SCC (Figures 5 and 6) indicates that SCC can be used as a simple surrogate for degree of mastitis. Accepting SCC as a continuous estimate of degree of mastitis allows the performance of the model to be evaluated on data from all time-points in lactation and not just the windows around mastitis treatments. Therefore, the reliability of the model was evaluated by examining the residuals from a regression of MR on log SCC (including adjustments for milk yield). These residuals can be interpreted as being the size of the error in MR (the assumptions involved are explored in the Discussion). The residual MR is shown in Figure 7.

The standard deviation of the residuals was 0.25. Of 68,032 observations, 1,457 had a residual >0.7; that is, 2.1% of the MR values deviated from the value expected from SCC measurements by more than +0.7. This result generated from the full lactation data of all cows in this study compared well with the likelihood of an MR of 0.7 belonging to the healthy population (1.7%) generated from the subset of true mastitic and true healthy cows in the time window around treatment (see Figure 5). The likelihood of a MR of 0.7 belonging to the healthy population of 1.7% implies that in a population of healthy cows (with the mean and SD of MR given above), 1.7% of the MR values generated by the model would be >0.7. Assuming that the true degree of masti-

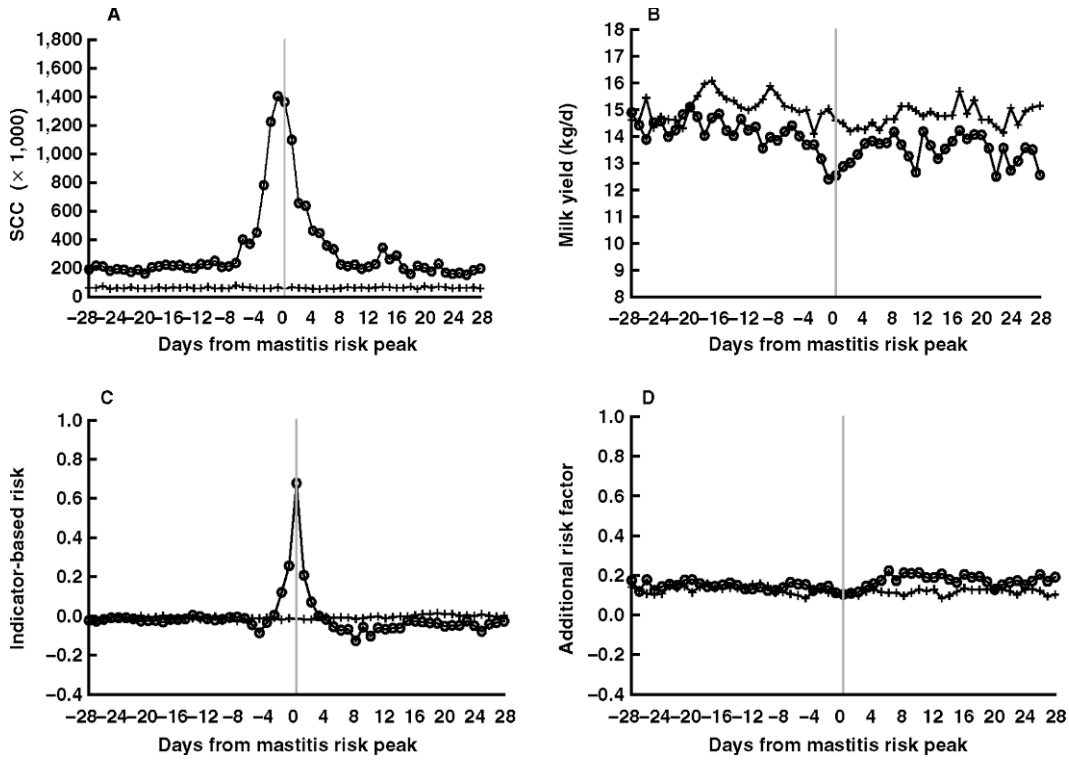

Figure 6. Time profiles of A) SCC, B) milk yield, C) indicator-based risk, and D) additional risk factor for true mastitic (—○—) and true healthy (—+—) cows. Profiles are time-aligned to the mastitis risk peak. The indicator-based risk was generated from changes in the level and slope of milk lactate dehydrogenase amount. The additional risk factor was generated from other animal- and herd-related factors available on-farm (DIM, udder characteristics, milking duration, and disease history). True healthy cows had no veterinary treatment and no peak in SCC; true mastitic cows had a veterinary treatment and a peak in SCC.

tis of the healthy population is zero, the model will have generated erroneous values above 0.7 for 1.7% of the healthy population. Similarly, the size of the residuals (model-generated MR – MR estimated from SCC) from the full data set can be interpreted under the assumption that SCC provides an error-free measure of true degree of mastitis (this assumption is examined in the Discussion) and that the regression model did not introduce any bias. Given this assumption, the proportion of residuals greater than +0.7 gives the proportion of occurrences of MR wrongly inflating the true MR by 0.7. Assuming that this proportion, found to be 2.1%, is the same across the range of true degrees of MR, then for healthy cows (i.e., true MR = 0), this is the proportion of records with erroneous model-generated MR values >0.7.

Using the assumptions given above, the percentage of residuals >0.7 provides an estimate of the proportion of healthy observations that would be misclassified as mastitis when using a classification threshold of 0.7. This equates to a specificity (the ability to avoid misclassifying healthy observations as mastitis) of (100 – 2.1) = 97.9%. Conversely, the percentage of residuals less than –0.3 was found to be 7.2%. This provides an estimate of the proportion of true mastitic observations; that is, the true degree of mastitis = 1, that would erroneously be assigned a model-generated degree of mastitis less than 0.7 and thus be misclassified as healthy. Assuming that mastitic cows have a MR of 1, this is equivalent to a sensitivity (the ability to avoid misclassifying mastitis observations as healthy) of (100 – 7.2) = 92.8%. Both these estimates of specificity and sensitivity are for a classifying threshold of 0.7.

Because the residuals are generated from all stages of lactation and degrees of mastitis, as indexed by SCC, they provide further information about the performance of the model being tested. The standard deviation of the residuals was greatest in early lactation (SD = 0.34 for 1 to 50 DIM) and declined as lactation progressed (SD = 0.20 for 100 to 150 DIM; Figure 7). There was also a tendency for the residuals to be greater for greater values of average MR, so the increased standard deviation of residuals in early lactation could have been because the frequency of mastitis occurrence was greater in early lactation. Therefore, the residuals from true healthy cows were examined relative to stage of

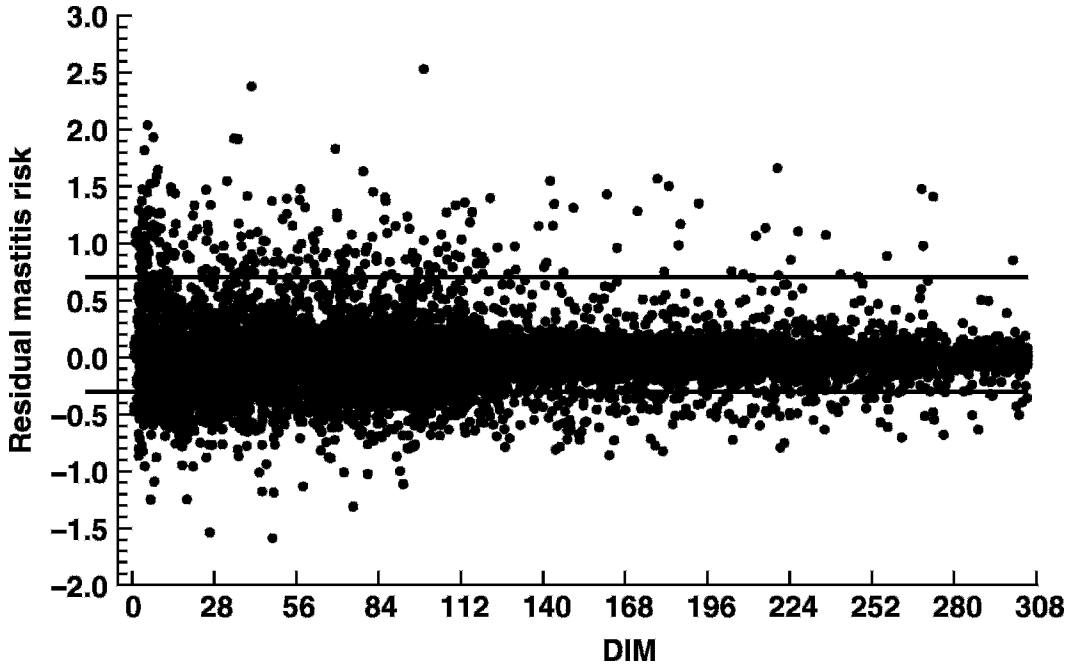

Figure 7. The residual mastitis risk (MR) relative to DIM. The residuals were derived from a regression of MR risk on SCC. The horizontal lines indicate residual MR levels of 0.7 and –0.3. Mastitis risk was generated by a model based on changes in the level and slope of milk lactate dehydrogenase amount.

lactation. There remained a greater error in early lactation: the percentages of errors >0.7 for the first, second, third, and fourth 50-d periods in lactation were 4.5, 1.9, 1.3, and 1.0, respectively.

DISCUSSION

The first aim of this study was to test a dynamic deterministic model for early identification of mastitis using LDH as an indicator in milk. To achieve this, a substantial set of milk samples was collected and analyzed for, among other things, LDH and SCC. The resulting data set, which included records of veterinary treatment, was large enough to allow a comprehensive evaluation of the main model output—the risk of acute mastitis. For interpreting these findings, it should be noted that this study was carried out on a farm with robotic milking. In conventional milking systems, mastitis surveillance is done on foremilk samples separately for each quarter, usually using the California Mastitis Test. Because of the practicalities of using a milking robot, we used SCC measured in proportional samples of composite milk taken throughout the milking. It has been shown that there is a strong relationship between quarter-level California Mastitis Test and composite milk SCC (Rasmussen et al., 2005), and we therefore judge that the findings of this study are relevant to other milking systems.

A highly significant difference in model-derived MR was found between true mastitic (i.e., received veterinary treatment for mastitis and had an SCC peak) and true healthy cows (i.e., no veterinary treatment and no positive systematic deviation in SCC to that point in lactation). The differences between true mastitic and true healthy cows were significant from –4 to +2 d realtive to veterinary treatment, indicating that one advantage of having a time-series of measurements is that mastitis can reliably be detected early. This would be a benefit of an automated, inline mastitis detection system based on an indicator such as LDH.

The second aim of this study was to use a test logic that allowed explicit examination of the time-related changes and a progressive scale of predicted mastitis state. We have previously reported favorable specificity/sensitivity values for this model (Chagunda et al., 2006). These values are, however, based on the use of a threshold to classify MR into positive and negative indications of mastitis in a fixed time window. In the present study, we have instead chosen to retain the time-profile of MR; that is, to retain the continuous nature of the model outputs. This is because we believe that there is valuable information in the time profiles that is lost when these are condensed into single specificity/sensitivity estimates. Using time profiles raises the issue of the appropriate reference time-point by which to align the profiles. We started with the time

of veterinary treatment as the reference point, which resulted in a median profile of MR that had a relatively broad peak that, although significantly different from true healthy controls, was relatively low (0.2; Figure 3). This timescale is relevant for considering the benefits of these time profiles relative to time of veterinary treatment (see above) but it does not reflect the true shape of the MR profiles of mastitic cows. There was a substantial variation in the time between peak MR and time of veterinary treatment (SD ± 2.8 d) and thus, a flattening of the median time profile when expressed relative to time of treatment. When the profiles were aligned relative to the peak MR, a much higher and narrower peak was found in the median profile. This reflects the shape of the individual MR peaks in this study. It has been documented that different mastitis pathogens develop into clinical mastitis at different rates and thus it is expected that different pathogens would generate different MR profiles (de Haas et al., 2002). In the present study, there were insufficient cases to allow meaningful results from splitting the cases into different pathogen types. This aspect remains to be explored.

As can be seen from Figure 5, the reliability of using a given value of MR to indicate mastitis decreases with decreasing MR value. This reliability is given as the likelihood of that value belonging to the distribution of healthy values, where healthy is defined as never having had a veterinary treatment of mastitis and never having had a systematic increase in logSCC above 2.25. The relationship between the likelihood of a MR belonging to the healthy population and the observed MR was readily approximated by the following exponential function (coefficients derived by log-linear regression): likelihood = $0.55 \times [\exp(-5 \times \text{MR})]$. Thus, for example, the likelihood of a MR of 0.7 belonging to the healthy population was estimated by this function to be 1.7%. Another way to assess the reliability of the model, by quantifying the deviations in MR values relative to the degree of mastitis, is discussed below.

When testing models, it should be remembered that the test is only as good as the reference values used. Clearly, if the reference value is wrong (e.g., a misdiagnosis of mastitis), then this will be counted as a failure of the model. The extent to which poor reference data compromises the test of a model increases the closer the true model performance is to the performance of the reference method. If the model is better than the reference method, then the "polarity" of the test is effectively reversed such that it more valid to view the test as being a test of the reference method relative to the model, not vice versa.

Although a number of statistical techniques exist for dealing with variable quality reference data (Wang et al., 2006), the ways in which the quality of the reference data and the test method used compromise the evaluation of a given model can be difficult to ascertain. In the present study, by using a graphically based, stepwise approach, it became clear that the quality of the reference data was affecting the test results. Supplementing the veterinary treatment records with SCC data increased the difference in MR between mastitic and healthy cows. Likewise, considering the time-related aspects; that is, recognizing that the time lag between peak infection and treatment is subject to random noise affected the results. Adjusting for this further increased the difference in MR between mastitic and healthy cows. Finally, the robustness of the test is affected by the assumed nature of the reference data. The shape of the time profiles of MR and SCC (Figures 5 and 6) provided good evidence that mastitis is not a binary condition but rather involves a continuous development from healthy to clinical mastitis through intermediate and increasing degrees of mastitis. Others have come to the same point of view (Green et al., 2004; Detilleux et al., 2006) and therefore we felt justified in using a continuous reference measure of degree of mastitis (or risk of developing clinical mastitis).

If it is accepted that mastitis is best described by degree of mastitis, then the robustness of any test based on specificity/sensitivity calculations is limited. To calculate specificity and sensitivity it is necessary to exclude a portion of the data if the underlying condition is continuous. As can be seen in Figure 8, the specificity (and sensitivity) of a detection method will always be poor when the true degree of mastitis is close to the classification threshold. Typically, the midrange values (the so-called subclinical cases) are excluded from calculations of specificity and sensitivity when the underlying condition is approximately continuous. Thus, a substantial and highly selective data reduction has to be undertaken to obtain useful specificity/sensitivity estimates. In the present study there were 58 true mastitic and 71 true healthy cases; these could provide the basis for a "sensible" specificity/sensitivity calculation but represented less than 1% of full data set. If more than 95% of your test data are excluded, then what is the worth of that test? Further, specificity/sensitivity calculations require that a threshold MR be defined for classifying cows as healthy or mastitic, and the resulting values of specificity and sensitivity will vary according to the threshold chosen (Norberg et al., 2004; Wang et al., 2006; Cavero et al., 2007). If instead, it can be demonstrated that a suitable measure is available to provide a continuous reference measure of degree of mastitis, then the model can be tested on the whole data set. In this situation, we suggest that it is more informative to describe the ability of the model to esti-

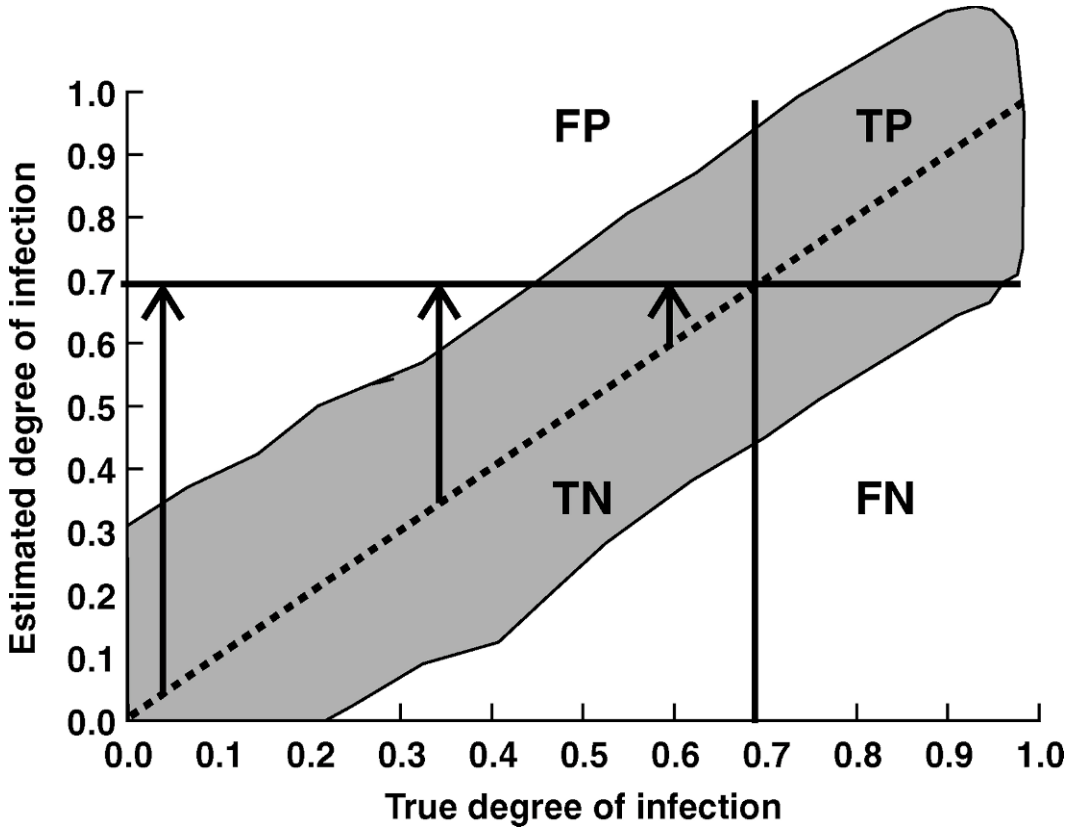

Figure 8. A schematic illustration of the inappropriateness of specificity/sensitivity calculations when the condition being measured is on a continuous scale. The shaded area denotes the observed relationship between true degree of infection and the degree of infection estimated by a given detection system; in this example, the error of estimation is constant over the range of degree of infection. Using a classification threshold of 0.7 for specificity/sensitivity calculations, estimates can be classified as true negatives (TN), false negatives (FN), true positives (TP), and false positives (FP). As the true degree of infection approaches the threshold, the proportion of estimates judged to be false increases dramatically. This does not reflect the fact that the error of estimation is constant over the full range of true degree of infection. The arrows denote the size of the error necessary for a given estimate to be misclassified (as a false positive); this clearly depends upon the true degree of infection.

mate degree of mastitis (in this study, MR) in terms of the residuals relative to the assumed true degree of mastitis. The estimates of reliability derived from the full data set using SCC as the degree of mastitis reference agreed well with those derived from quantifying the differences between true mastitic and true healthy cases, and are comparable with the highest reported values from other mastitis detection studies (Hillerton, 2000; de Mol and Ouweltjes, 2001; Cavero et al., 2007).

Two important assumptions are made in using SCC as the degree of mastitis reference: first, that SCC is an adequate measure of true degree of mastitis, and second, because SCC is the reference measure, that it is free from measurement error. Both of these assumptions are open to question and warrant further examination. In the Materials and Methods section, "reasonable certainty" was the term used for describing the classification of observations and cows as mastitic or healthy. This indicates that there is no absolute, undisputed gold standard for defining mastitis state even when veterinary records are supplemented with SCC data. The same applies to other mastitis measurements, including bacteriological counts. This is particularly the case when the desired definition of mastitis state is the degree of mastitis rather than the binary classification of healthy or mastitic. This lack of gold standard has been discussed by others (Dohoo and Leslie, 1991; Sloth et al., 2003), and attempts to achieve consensus have been made (Rasmussen et al., 2005). It seems that the best reference for true degree of mastitis will be a combination of the available mastitis measures. A number of statistical techniques exist that can address this problem of inadequate reference measures (Toft et al., 2005; Wang et al., 2006) although to date they have not been widely used within animal science.

By combining these with data filtering and smoothing techniques, it should be possible to evaluate which combination of mastitis measures best indexes degree of

mastitis while at the same time dealing with random noise in these measures. We consider this the logical next step in the development of the test rationale for explicit examination of time-related changes in degree of mastitis, and thus an issue worth pursuing.

CONCLUSIONS

A dynamic deterministic model for early identification of mastitis using LDH as an indicator in milk was tested and found to be as accurate as the best published mastitis detection systems. It was able to detect significant differences between mastitic and healthy cows 4 d before treatment, indicating that one advantage of having a time-series of measurements is that mastitis can reliably be detected early. It was also found that specificity/sensitivity calculations are inappropriate for examination of the time-related changes and for evaluating a progressive scale of predicted mastitis state.

ACKNOWLEDGMENTS

We gratefully acknowledge the valuable contribution of the farm staff at the Danish Cattle Research Centre (Tjele, Denmark), Jens Clausen, Carsten Berthelsen, Peter Løvendahl, Jes Nielsen, and Connie Middelhede for their efforts in securing this massive milk data set. This study, which was part of the Biosens project, was funded by the Danish Ministry of Food, Agriculture and Fisheries and the Danish Cattle Association.

REFERENCES

Cavero, D., K. H. Tölle, C. Buxadé, and J. Krieter. 2007. Mastitis detection in dairy cows by application of fuzzy logic. Livest. Sci. 105:207–213.

Chagunda, M. G. G., N. C. Friggens, T. Larsen, and M. D. Rasmussen. 2006. A biological model for detecting individual cow mastitis risk based on lactate dehydrogenase: Model description. J. Dairy Sci. 89:2980–2998.

de Haas, Y., H. W. Barkema, and R. F. Veerkamp. 2002. The effect of pathogen-specific clinical mastitis on the lactation curve for somatic cell count. J. Dairy Sci. 85:1314–1323.

de Mol, R. M., and W. Ouweltjes. 2001. Detection model for mastitis in cows milked in an automatic milking system. Prev. Vet. Med. 49:71–82.

Detilleux, J., F. Vangroenweghe, and C. Burvenich. 2006. Mathematical model of the acute inflammatory response to *Escherichia coli* in intramammary challenge. J. Dairy Sci. 89:3455–3465.

Dohoo, I. R., and K. E. Leslie. 1991. Evaluation of changes in somatic cell counts as indicators of new intramammary infections. Prev. Vet. Med. 10:225–237.

Gardner, I. A., and M. Greiner. 2006. Receiver-operating characteristic curves and likelihood ratios: Improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. Vet. Clin. Pathol. 35:8–17.

Green, M. J., P. R. Burton, L. E. Green, Y. H. Schukken, A. J. Bradley, E. J. Peeler, and G. F. Medley. 2004. The use of Markov chain Monte Carlo for analysis of correlated binary data: Patterns of somatic cells in milk and the risk of clinical mastitis in dairy cows. Prev. Vet. Med. 64:157–174.

Greiner, M., and I. A. Gardner. 2000. Epidemiologic issues in the validation of veterinary diagnostic tests. Prev. Vet. Med. 45:3–22.

Hillerton, J. E. 2000. Detecting mastitis cow-side. Pages 48–53 in Proc. 39th Annual Meeting National Mastitis Council, National Mastitis Council, Madison, WI.

IDF. 1997. Recommendations for presentation of mastitis-related data. Bulletin 321. International Dairy Federation, Brussels, Belgium.

Larsen, T. 2005. Determination of lactate dehydrogenase (LDH) activity in milk by a fluorometric assay. J. Dairy Res. 72:209–216.

Lasko, T. A., J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado. 2005. The use of receiver operating characteristic curves in biomedical informatics. J. Biomed. Inform. 38:404–415.

Norberg, E., H. Hogeveen, I. R. Korsgaard, N. C. Friggens, K. H. M. N. Sloth, and P. L. Løvendahl. 2004. Electrical conductivity in milk: Ability to predict mastitis status. J. Dairy Sci. 87:1099–1107.

Rasmussen, M. D., M. Bjerring, and F. Skjøth. 2005. Visual appearance and CMT score of foremilk of individual quarters in relation to cell count of cows milked automatically. J. Dairy Res. 72:49–56.

Sloth, K. H. M. N., N. C. Friggens, P. L. Løvendahl, P. H. Andersen, J. Jensen, and K. L. Ingvartsen. 2003. Potential for improving description of bovine udder health status by combined analysis of milk parameters. J. Dairy Sci. 86:1221–1232.

Toft, N., E. Jørgensen, and S. Højsgaard. 2005. Diagnosing diagnostic tests: Evaluating the assumptions underlying the estimation of sensitivity and specificity in the absence of a gold standard. Prev. Vet. Med. 68:19–33.

Wang, C., B. W. Turnbull, Y. T. Gröhn, and S. S. Nielsen. 2006. Estimating receiver operating characteristic curves with covariates when there is no perfect reference test for diagnosis of Johne's disease. J. Dairy Sci. 89:3038–3046.