

Mastitis in Dairy Production: Estimation of Sensitivity, Specificity and Disease Prevalence in the Absence of a Gold Standard

Caroline UHLER

Mastitis, a worldwide endemic disease of dairy cows, is an important cause of decreased efficiency in milk production. Early medical treatment can reduce the nonreversible losses in milk production caused by this infection. Various diagnostic tests for mastitis are available, including a test measuring the electrical conductivity of milk (MEC test), the industry standard of somatic cell counting (SCC test), a bacteriological test, and a recently developed test measuring mammary associated amyloid A (MAA test). None of these tests is considered a gold standard, however. The aim of the present study was to determine which of these tests provides the best results, and at what cost, to improve the efficiency of milk production. For this study, 25 cows were tested at all four quarters of the udder with each of the aforementioned mastitis diagnostic tests. Based on the data, the disease prevalence as well as the sensitivity and the specificity of the four tests were estimated with a Bayesian approach by extending the Hui and Walter model with two independent tests and two populations to a model with four partially dependent tests and one population. This model was further combined with a receiver operating characteristics analysis to estimate the overall test accuracy.

Key Words: Bayesian approach; Bayesian model averaging; Hui and Walter model; Mastitis diagnostic tests; MCMC; RJMCMC; ROC curve.

1. INTRODUCTION

Defined as any inflammatory process in the mammary gland (International Dairy Federation 1987), mastitis occurs in two different forms: clinical and subclinical. Clinical mastitis involves clinical signs in the udder (e.g., reddening, swelling, pain, high temperature) or visible changes in the milk (e.g., flakes, clots). The more important form, subclinical mastitis, can be detected only by laboratory analysis of parameters related to the inflammatory process, such as somatic cell count. The present study focused on subclinical mastitis.

Caroline Uhler was a M.S. Student at the Institute of Mathematics, University of Zurich, 8057 Zurich, Switzerland, and is now at the Department of Statistics, University of California, Berkeley, CA 94704, USA (E-mail: cuhler@stat.berkeley.edu).

© 2009 American Statistical Association and the International Biometric Society
Journal of Agricultural, Biological, and Environmental Statistics, Volume 14, Number 1, Pages 79–98
DOI: 10.1198/jabes.2009.0005

The udder is divided into four biologically independent quarters, and mastitis is associated with one of these quarters, not with the entire cow. Pathogens usually can be found in infected quarters. Mastitis pathogens can be environmental pathogens, often associated with unhygienic environmental conditions. In addition, some obligate contagious mastitis pathogens or commensals can be spread from cow to cow through unhygienic milking practices. The udder is naturally protected from bacterial entry, but in cows bred for easier milking, this natural protection is decreased, and bacteria can enter the udder more easily, increasing the risk of mastitis.

In this study we investigated the diagnostic efficiency of four mastitis diagnostic tests: a bacteriological test, somatic cell count (SCC), measurement of mammary-associated amyloid A (MAA), and measurement of electrical conductivity (MEC). We describe these tests briefly here; more details have been provided by Whyte et al. (2004) and Hogeveen (2005).

BACTERIOLOGICAL TEST

In most cases, mastitis is caused by such bacteria as streptococci, staphylococci, coagulase-negative staphylococci (CNS), and corynebacteria. The different pathogens require different antibiotics to achieve quick healing; thus, a bacteriological test may be important. This test is a qualitative test; an udder quarter is considered to test positive if the foregoing bacteria are found and negative otherwise. For the bacteriological test, contamination from external sources can cause false-positive results, and errors in the agars can inhibit the growth of bacteria and lead to false-negative results. In addition, the role of corynebacteria and CNS as udder pathogens is not clear, providing another potential source of false-positive test results. This test is the costliest of the four tests, at about 12 Swiss francs (CHF) per quarter.

SOMATIC CELL COUNT

This is the most widely used test for detecting mastitis and is very important in herd management. The cells found in the milk are somatic cells for defense against the bacteria and stem from the blood. SCC is measured in cells per milliliter. An infection is assumed when a specific concentration (normally 100,000 cells/mL of milk) is exceeded. Both false-positive and false-negative results can occur on the farm, from incorrect attribution of samples due to improper storage or transport, and in the laboratory due to sample misidentification, errors in counting (including improper calibration), or carry-over of cells from a sample with high cell counts. The cost of this test is about 1.5 CHF per quarter.

MAMMARY-ASSOCIATED AMYLOID A

The initial response of the mammalian immune system to an infection is the production of acute-phase proteins, which trigger the body's defense and repair mechanisms. An acute-phase protein, MAA is produced in the udder and helps trigger the cow's immune system to respond to an antigen. Usually, an infection is assumed when the concentration of MAA exceeds 400 ng/mL. The production of MAA can represent a response to a bacterial infection, but may also result from physical damage or a stress situation, explaining the occurrence of false-positive test results. In addition, both false-positive and

false-negative results can result from measurement and dilution errors. This test costs approximately 1.5 CHF per quarter.

MEASUREMENT OF ELECTRICAL CONDUCTIVITY

Measuring the electrical conductivity of milk to detect mastitis dates back to the first half of the twentieth century. An infection of the udder causes tissue damage and increases the electrical conductivity of the milk. Usually, an infection is assumed when the electrical conductivity exceeds 5.5 milliSiemens. The conductivity can be affected by tissue damage due to infection as well as by many other factors, including the course of lactation, milking intervals, milk fat content, milk temperature, and foods ingested, possibly leading to false-positive and false-negative results. The MEC test has the lowest cost, about 0.10 CHF per quarter.

To get an idea of the damage caused by mastitis, Professor P. Rüschi of the University of Zurich has calculated an average annual cost of 350 CHF per cow for Swiss farmers (Walkenhorst 2004). This amount includes 60 CHF for veterinary and medical treatment and 50 CHF for premature replacement of cows weakened by mastitis. But the bulk of this cost, the remaining 240 CHF, is attributed to decreased milk production and discarded milk. Early treatment can decrease the nonreversible losses in milk production due to mastitis. Therefore, tests with a high probability of detecting mastitis when present (i.e., high sensitivity) and of providing a negative result in noninfected cows (i.e., high specificity) are needed. It would be simple to estimate the sensitivity and specificity of these tests if a perfect test were available for comparison; however, a gold standard for the detection of mastitis has not yet been established.

For estimating test accuracy in the absence of a gold standard, each test is customarily evaluated against other tests with their own errors by applying the tests simultaneously to each individual. Hui and Walter (1980) considered the case where two tests (both with unknown sensitivity and specificity) were simultaneously applied to individuals from two populations with differing disease prevalences. Assuming conditional independence of the tests, they showed how the sensitivity and specificity of both tests, as well as the prevalence in both populations, can be estimated by the maximum likelihood method. Since then, several other approaches have been developed for evaluating tests in the Hui and Walter model. (See Enøe, Georgiadis, and Johnson 2000 for a review of existing methods, including the Bayesian approach.) In the present work we used the Markov chain Monte Carlo (MCMC) methodology and performed the computations by Gibbs sampling. Gilks, Richardson, and Spiegelhalter (1995) have provided a thorough introduction to MCMC methods and their applications. Recently, Toft et al. (2007) discussed different tools for assessing convergence of MCMC methods based on the Bayesian approach to the Hui and Walter model.

The Hui and Walter model assumes conditional independence of the two tests; however, it is not uncommon to have data from conditionally dependent tests. Black and Craig (2002) extended the Hui and Walter model using MCMC methodology to deal with such data. They developed four models that vary in terms of the form of dependence and used reversible-jump MCMC (RJ-MCMC) to move from one model to another. Recently, Toft,

Jørgensen, and Højsgaard (2005) reviewed the importance and implications of the assumptions of the Hui and Walter model. They criticized the work of Black and Craig based on the fact that in their model the number of parameters exceeds the degrees of freedom. This results in a lack of identifiability and implies that estimates cannot be reliably obtained from the data.

To estimate the disease prevalence as well as the sensitivities and specificities of the four mastitis diagnostic tests, in this work we extended the model introduced by Black and Craig to a model with one population and four tests. But to get an identifiable problem, we allowed at most pairwise dependence of the tests. From a biological standpoint, this should be sufficient, because only the SCC and MAA tests measure similar biological processes and thus could reasonably be expected to be dependent. Moreover, the model was combined with a receiver operating characteristics (ROC) analysis to estimate the overall accuracy of each test. Using these estimates, the optimal test for mastitis can be found, allowing earlier detection of mastitis to increase the efficiency of milk production.

2. MODEL

2.1 NOTATION

When applying four tests to one population, the observed data can be classified into a $2 \times 2 \times 2 \times 2$ contingency table. Each cell holds the count of tested individuals with a given combination of the four binary tests, T_j , $j = 1, 2, 3, 4$, where T_1 denotes the bacteriological test, T_2 denotes the SCC test, T_3 denotes the MAA test, and T_4 denotes the MEC test. A negative result on the j th test is denoted by $T_j = 0$; a positive result, by $T_j = 1$.

The parameters of primary interest in diagnostic testing are disease prevalence, denoted by π , and test sensitivity and specificity, denoted by Se_j and Sp_j , $j = 1, 2, 3, 4$. Let D denote the truly diseased status and \bar{D} denote the nondiseased status. Prevalence is defined as the probability of being truly diseased, that is, $\pi = P(D)$. The sensitivity of the j th test measures its ability to detect a disease when it is present, that is, $Se_j = P(T_j = 1|D)$. The specificity of the j th test measures its ability to provide a negative result in noninfected individuals, that is, $Sp_j = P(T_j = 0|\bar{D})$.

If the true disease status of each individual were known, then each cell count n_{ijkl} of the $2 \times 2 \times 2 \times 2$ contingency table could be broken up into $n_{ijkl} = z_{ijkl} + y_{ijkl}$, where z_{ijkl} is the number of truly diseased individuals and y_{ijkl} the number of nondiseased individuals in a given cell. The outcome of the bacteriological test is denoted by i , the outcome of the SCC test is denoted by j , the outcome of the MAA test is denoted by k , and the outcome of the MEC test is denoted by l . Instead of a single $2 \times 2 \times 2 \times 2$ table, the data then can be summarized in two separate tables, one for each disease status. The estimates of disease prevalence (π), sensitivity (Se_j , $j = 1, 2, 3, 4$), and specificity (Sp_j , $j = 1, 2, 3, 4$) then can be easily calculated using the following formulas, where Z denotes the total number of disease-positive individuals, Y denotes the total number of disease-negative individuals, N denotes the total number of individuals, and the dot subscripts indicate the sum over that

index:

$$\begin{aligned} \hat{\pi} &= \frac{Z}{N}, \\ \hat{Se}_1 &= \frac{z_{1\dots}}{Z}, & \hat{Se}_2 &= \frac{z_{\cdot 1\dots}}{Z}, & \hat{Se}_3 &= \frac{z_{\dots 1}}{Z}, & \hat{Se}_4 &= \frac{z_{\dots 1}}{Z}, \\ \hat{Sp}_1 &= \frac{y_{0\dots}}{Y}, & \hat{Sp}_2 &= \frac{y_{\cdot 0\dots}}{Y}, & \hat{Sp}_3 &= \frac{y_{\dots 0}}{Y}, & \hat{Sp}_4 &= \frac{y_{\dots 0}}{Y}. \end{aligned}$$

Unlike the situation considered in the last paragraph, in the data available for this study the true disease status of each individual is unknown. Using Bayesian inference, these unknown counts can be simulated to get posterior distributions and estimates of the parameters. The algorithm for doing this is given in Section 2.3.

2.2 DEPENDENCE STRUCTURE

As defined and discussed in more detail by Gardner et al. (2000), two tests are *conditionally independent* when the sensitivity (or specificity) of the second test does not depend on whether the results of the first test are positive or negative in infected (or noninfected) individuals. This means that $P(T_2 = 1|T_1 = 1, D) = P(T_2 = 1|T_1 = 0, D) = P(T_2 = 1|D)$.

Tests that measure similar biological processes are likely to be positively dependent when conditioning on the true disease status. For example, two different tests measuring the serum antibody responses to infectious agents will tend to follow a similar time-dependent pattern. False-negative results on both tests are more likely early in the course of infection, when the concentration of antibodies is lower. In addition, false-positive serologic responses due to vaccination or cross-reacting antibodies tend to be positively correlated on different serologic tests.

Somatic cells and MAA are both produced by the body's defense mechanisms, which as explained earlier, could lead to a positive dependence between the SCC and the MAA tests. There is no apparent biological relationship between other pairs of tests. This information is considered in the model. The positive conditional dependence between the sensitivity of the SCC test and the MAA test can be expressed as

$$p_{\cdot 11 \cdot | D} := P(T_2 = 1, T_3 = 1|D) > P(T_2 = 1|D)P(T_3 = 1|D),$$

and the dependence of test specificities can be expressed as

$$p_{\cdot 00 \cdot | \bar{D}} := P(T_2 = 0, T_3 = 0|\bar{D}) > P(T_2 = 0|\bar{D})P(T_3 = 0|\bar{D}).$$

Note that a dependence of test sensitivities does not necessarily imply a dependence of test specificities and vice versa.

Thus the full conditional dependence model, where test responses are correlated for both disease states, introduces two additional parameters $p_{\cdot 11 \cdot | D}$ and $p_{\cdot 00 \cdot | \bar{D}}$ with the constraints $p_{\cdot 11 \cdot | D} > Se_2Se_3$ and $p_{\cdot 00 \cdot | \bar{D}} > Sp_2Sp_3$, whereas the partial dependence models include only one of these parameters. The model that includes only $p_{\cdot 11 \cdot | D}$ is called the *disease-positive model*, and the model with $p_{\cdot 00 \cdot | \bar{D}}$ is called the *disease-negative model*. In the independence model, $p_{\cdot 11 \cdot | D} = Se_2Se_3$, and $p_{\cdot 00 \cdot | \bar{D}} = Sp_2Sp_3$.

For each disease status D and \bar{D} , the probability of being classified into each of the 16 cells of the contingency tables is denoted by $p_{ijkl|D}$ and $p_{ijkl|\bar{D}}$. These classification

Table 1. Formulas for the probabilities $p_{ijkl|D}$.

$Se_2 Se_3 \leq p_{\cdot 11 \cdot D} \leq \min(Se_2, Se_3)$	
$p_{0000 D} = (1 - Se_1)(1 - Se_2 - Se_3 + p_{\cdot 11 \cdot D})(1 - Se_4)$	
$p_{1000 D} = Se_1(1 - Se_2 - Se_3 + p_{\cdot 11 \cdot D})(1 - Se_4)$	
$p_{0100 D} = (1 - Se_1)(Se_2 - p_{\cdot 11 \cdot D})(1 - Se_4)$	
$p_{1100 D} = Se_1(Se_2 - p_{\cdot 11 \cdot D})(1 - Se_4)$	
$p_{0010 D} = (1 - Se_1)(Se_3 - p_{\cdot 11 \cdot D})(1 - Se_4)$	
$p_{1010 D} = Se_1(Se_3 - p_{\cdot 11 \cdot D})(1 - Se_4)$	
$p_{0110 D} = (1 - Se_1)p_{\cdot 11 \cdot D}(1 - Se_4)$	
$p_{1110 D} = Se_1 p_{\cdot 11 \cdot D}(1 - Se_4)$	
$p_{0001 D} = (1 - Se_1)(1 - Se_2 - Se_3 + p_{\cdot 11 \cdot D})Se_4$	
$p_{1001 D} = Se_1(1 - Se_2 - Se_3 + p_{\cdot 11 \cdot D})Se_4$	
$p_{0101 D} = (1 - Se_1)(Se_2 - p_{\cdot 11 \cdot D})Se_4$	
$p_{1101 D} = Se_1(Se_2 - p_{\cdot 11 \cdot D})Se_4$	
$p_{0011 D} = (1 - Se_1)(Se_3 - p_{\cdot 11 \cdot D})Se_4$	
$p_{1011 D} = Se_1(Se_3 - p_{\cdot 11 \cdot D})Se_4$	
$p_{0111 D} = (1 - Se_1)p_{\cdot 11 \cdot D}Se_4$	
$p_{1111 D} = Se_1 p_{\cdot 11 \cdot D}Se_4$	

probabilities are functions of the sensitivities and specificities of the four tests and of the parameters $p_{\cdot 11 \cdot |D}$ and $p_{\cdot 00 \cdot |D}$. The formulas for the probabilities $p_{ijkl|D}$ are given in Table 1. The probabilities $p_{ijkl|\bar{D}}$ are computed analogously.

To sum up, the structure of this model results in 11 parameters: one parameter for the prevalence, four parameters each for the sensitivities and the specificities, and two parameters for the dependence structure. But four tests and one population provide 15 degrees of freedom, so the number of parameters does not exceed the number of degrees of freedom and thus meets the requirements stipulated by [Toft, Jørgensen, and Højsgaard \(2005\)](#).

2.3 ALGORITHM

As in the model of [Black and Craig \(2002\)](#), in the present study we used a combination of Gibbs sampling and Metropolis–Hastings to estimate the parameters. Initially, starting values for the 11 parameters must be specified. Starting values for each parameter are sampled from their prior distribution. We chose the beta distribution as the prior for the sensitivities and specificities of the four tests, as well as for the prevalence (see Table 2). Using beta distributions as the prior greatly simplifies calculations; moreover, beta distributions can yield a large array of potential shapes. For the full conditional dependence model, we must also define priors for the other two parameters, $p_{\cdot 11 \cdot |D}$ and $p_{\cdot 00 \cdot |D}$. Little a priori knowledge of the degree of dependence is assumed, and thus a uniform distribution

Table 2. Prior distribution for each model quantity.

$\pi \sim \text{Beta}(a_\pi, b_\pi)$	
$Se_j \sim \text{Beta}(a_{Se_j}, b_{Se_j}), j = 1, 2, 3, 4$	$Sp_j \sim \text{Beta}(a_{Sp_j}, b_{Sp_j}), j = 1, 2, 3, 4$
$p_{\cdot 11 \cdot D} \sim \text{Uniform}(Se_2 Se_3, \min(Se_2, Se_3))$	$p_{\cdot 00 \cdot D} \sim \text{Uniform}(Sp_2 Sp_3, \min(Sp_2, Sp_3))$

is used. With this choice, the joint probability is equally likely to be anywhere between its values for independence and those for complete positive dependence.

Given the observed counts n_{ijkl} , the missing counts of the truly diseased individuals consist of independently distributed binomial variates with

$$z_{ijkl}|n_{ijkl} \sim \text{Bin}(n_{ijkl}, P_{ijkl}),$$

where P_{ijkl} is the probability of being diseased given that the individual has test combination (i, j, k, l) . Using the Bayes theorem leads to the following equation, through which the probabilities P_{ijkl} can be computed by substituting the formulas given in Table 1:

$$\begin{aligned} P_{ijkl} &= P(D|T_1 = i, T_2 = j, T_3 = k, T_4 = l) \\ &= P(T_1 = i, T_2 = j, T_3 = k, T_4 = l|D)P(D) \\ &\quad / (P(T_1 = i, T_2 = j, T_3 = k, T_4 = l|D)P(D) \\ &\quad + P(T_1 = i, T_2 = j, T_3 = k, T_4 = l|\bar{D})P(\bar{D})) \\ &= \frac{p_{ijkl|D}\pi}{p_{ijkl|D}\pi + p_{ijkl|\bar{D}}(1 - \pi)}. \end{aligned}$$

2.3.1 Independence Model

In this model, new values for the prevalence, the sensitivities and the specificities can be sampled directly from their full conditional distributions with the foregoing computed z_{ijkl} values. The full conditional distributions are given in Table 3.

New values for the prevalence, sensitivity, and specificity are sampled independently of one another from these distributions. With this step, the first iteration of the algorithm for the independence model is completed. The next iteration begins by computing the probabilities P_{ijkl} with the values of the model quantities from the preceding iteration. These probabilities are then substituted into the conditional binomial distributions, the missing z_{ijkl} 's are sampled, and, finally, new values for the model quantities are resampled from their full conditional distributions.

Table 3. Full conditional distributions of all model quantities in the independence model.

$\pi \sim \text{Beta}(\pi 1 + Z, 1 + N - Z)$
$\text{Se}_1 \sim \text{Beta}(\text{Se}_1 1 + z_{1\dots}, 1 + z_{0\dots})$
$\text{Se}_2 \sim \text{Beta}(\text{Se}_2 1 + z_{.1\dots}, 1 + z_{.0\dots})$
$\text{Se}_3 \sim \text{Beta}(\text{Se}_3 1 + z_{..1\dots}, 1 + z_{..0\dots})$
$\text{Se}_4 \sim \text{Beta}(\text{Se}_4 1 + z_{\dots 1}, 1 + z_{\dots 0})$
$\text{Sp}_1 \sim \text{Beta}(\text{Sp}_1 1 + n_{0\dots} - z_{0\dots}, 1 + n_{1\dots} - z_{1\dots})$
$\text{Sp}_2 \sim \text{Beta}(\text{Sp}_2 1 + n_{.0\dots} - z_{.0\dots}, 1 + n_{.1\dots} - z_{.1\dots})$
$\text{Sp}_3 \sim \text{Beta}(\text{Sp}_3 1 + n_{..0\dots} - z_{..0\dots}, 1 + n_{..1\dots} - z_{..1\dots})$
$\text{Sp}_4 \sim \text{Beta}(\text{Sp}_4 1 + n_{\dots 0} - z_{\dots 0}, 1 + n_{\dots 1} - z_{\dots 1})$

2.3.2 Full Conditional Dependence Model

In the full conditional dependence model, a possible conditional dependence between the SCC test and the MAA test is assumed. Because of the parameter space restrictions introduced by the additional parameters $p_{.11.|D}$ and $p_{.00.|\bar{D}}$ in this model, using just the Gibbs sampler and sampling directly from the full conditionals is difficult. Therefore, a Metropolis–Hastings step for updating the parameters $(\text{Se}_2, \text{Se}_3, p_{.11.|D})$ and $(\text{Sp}_2, \text{Sp}_3, p_{.00.|\bar{D}})$ is introduced. Note that this Metropolis–Hastings step involves only the second and third tests.

If there were no restrictions on the probability vector $p_D := (p_{.00.|D}, p_{.10.|D}, p_{.01.|D}, p_{.11.|D})$ (resp. on the vector $p_{\bar{D}} := (p_{.00.|\bar{D}}, p_{.10.|\bar{D}}, p_{.01.|\bar{D}}, p_{.11.|\bar{D}})$), and a Dirichlet(1, 1, 1, 1) prior were used, then the full conditional distribution would be Dirichlet($z_{.00.} + 1, z_{.10.} + 1, z_{.01.} + 1, z_{.11.} + 1$) (resp. Dirichlet($n_{.00.} - z_{.00.} + 1, n_{.10.} - z_{.10.} + 1, n_{.01.} - z_{.01.} + 1, n_{.11.} - z_{.11.} + 1$)). These distributions are used as proposal distributions in the Metropolis–Hastings algorithm.

As explained by Black and Craig (2002), the first step is to generate the proposed probabilities $p_D^* := (p_{.00.|D}^*, p_{.01.|D}^*, p_{.10.|D}^*, p_{.11.|D}^*)$ from the full conditional distribution Dirichlet($z_{.00.} + 1, z_{.10.} + 1, z_{.01.} + 1, z_{.11.} + 1$). The next step is to compute the proposed sensitivities, $\text{Se}_2^* = p_{.11.|D}^* + p_{.10.|D}^*$ and $\text{Se}_3^* = p_{.11.|D}^* + p_{.01.|D}^*$. If $p_{.11.|D}^* > \text{Se}_2^* \text{Se}_3^*$, then move to the new proposed parameters with probability

$$P(\text{move}) = \min\left(1, \frac{\min(\text{Se}_2, \text{Se}_3) - \text{Se}_2 \text{Se}_3}{\min(\text{Se}_2^*, \text{Se}_3^*) - \text{Se}_2^* \text{Se}_3^*} \prod_{t=2}^3 \left(\frac{\text{Se}_t^*}{\text{Se}_t}\right)^{a_{\text{Se}_t} - 1} \left(\frac{1 - \text{Se}_t^*}{1 - \text{Se}_t}\right)^{b_{\text{Se}_t} - 1}\right);$$

otherwise, do not update the parameters in this iteration. Updating $p_{.00.|\bar{D}}$, Sp_2 and Sp_3 works analogously. Finally, generate $\text{Se}_1, \text{Se}_4, \text{Sp}_1, \text{Sp}_4$, and π from the full conditional distributions given in Table 3 as in the independence model.

Thus far, the independence and the full conditional dependence models have been constructed. Updating the partial dependence models is accomplished analogously. Because the dependence structure of the SCC and MAA tests is unclear, we want to combine all four models. We account for this uncertainty by allowing the Markov chain to jump between the four models. This procedure is explained in the next section and was discussed in more detail by Hoeting et al. (1999).

2.4 BAYESIAN MODEL AVERAGING

Let M_1 be the independence model, M_2 be the partially dependent disease positive model, M_3 be the partially dependent disease negative model and M_4 be the full conditional dependence model. Note that these expressions are related only to the dependence structure of the second and third tests. Thus the models are as follows:

$$\begin{aligned} M_1: & \quad p_{.11.|D} = \text{Se}_2 \text{Se}_3, & \quad p_{.00.|\bar{D}} = \text{Sp}_2 \text{Sp}_3, \\ M_2: & \quad p_{.11.|D} > \text{Se}_2 \text{Se}_3, & \quad p_{.00.|\bar{D}} = \text{Sp}_2 \text{Sp}_3, \\ M_3: & \quad p_{.11.|D} = \text{Se}_2 \text{Se}_3, & \quad p_{.00.|\bar{D}} > \text{Sp}_2 \text{Sp}_3, \\ M_4: & \quad p_{.11.|D} > \text{Se}_2 \text{Se}_3, & \quad p_{.00.|\bar{D}} > \text{Sp}_2 \text{Sp}_3. \end{aligned}$$

Next, we construct a Markov chain on these four models. Each model is a priori assumed to be equally likely, and again the Metropolis–Hastings algorithm is used.

- First, consider a move from M_1 to M_2 or from M_3 to M_4 . This involves switching from $p_{\cdot 11 \cdot | D} = \text{Se}_2 \text{Se}_3$ to $p_{\cdot 11 \cdot | D} > \text{Se}_2 \text{Se}_3$. Thus we sample γ from the uniform distribution $U(0, \min(\text{Se}_2, \text{Se}_3) - \text{Se}_2 \text{Se}_3)$ and define $p_{\cdot 11 \cdot | D}^* = \text{Se}_2 \text{Se}_3 + \gamma$. As explained by Gilks, Richardson, and Spiegelhalter (1995), the probability of a move from the current point θ^{t-1} to a candidate point θ^* in the Metropolis–Hastings algorithm is

$$\begin{aligned} P(\text{move}) &= \min\left(1, \frac{p(\theta^* | y) p(\theta^{t-1} | \theta^*)}{p(\theta^{t-1} | y) p(\theta^* | \theta^{t-1})}\right) \\ &= \min\left(1, \frac{p(y | \theta^*) p(\theta^*)}{p(y | \theta^{t-1}) p(\theta^{t-1})} \frac{p(\theta^{t-1} | \theta^*)}{p(\theta^* | \theta^{t-1})}\right), \end{aligned}$$

where y denotes the observed data. Thus the acceptance probability for a move from M_1 to M_2 or from M_3 to M_4 is

$$\begin{aligned} P(\text{move}) &= \min\left(1, \frac{p(n | M_2, \theta_2) p(\theta_2 | M_2) p(M_2)}{p(n | M_1, \theta_1) p(\theta_1 | M_1) p(M_1)} \frac{1}{(\min(\text{Se}_1, \text{Se}_2) - \text{Se}_1 \text{Se}_2)^{-1}}\right) \\ &= \min\left(1, \frac{P(z | \text{Se}_1, \text{Se}_2, p_{\cdot 11 \cdot | D})}{P(z | \text{Se}_1, \text{Se}_2)}\right), \end{aligned} \tag{2.1}$$

where $\theta_1 = \{\text{Se}_2, \text{Se}_3\}$, $\theta_2 = \{\text{Se}_2, \text{Se}_3, p_{\cdot 11 \cdot | D}\}$, and n and z denote the table n_{ijkl} and the table z_{ijkl} , where $i, j, k, l \in \{0, 1\}$. For the second equality, we assume that the prior probability of each model is $\frac{1}{4}$, that θ_1 and θ_2 affect only the truly diseased individuals, and that

$$p(\theta_2 | M_2) = p(p_{\cdot 11 \cdot | D} | \text{Se}_2, \text{Se}_3, M_2) p(\theta_1 | M_1).$$

- For the reverse moves from M_2 to M_1 or from M_4 to M_3 , define $p_{\cdot 11 \cdot | D}^* = \text{Se}_2 \text{Se}_3$. The acceptance probability is computed analogously to (2.1):

$$P(\text{move}) = \min\left(1, \frac{P(z | \text{Se}_2, \text{Se}_3)}{P(z | \text{Se}_2, \text{Se}_3, p_{\cdot 11 \cdot | D})}\right).$$

- Similar acceptance probabilities are used to move between M_1 and M_3 or between M_2 and M_4 .

Therefore, one iteration of the algorithm comprises first updating the parameters, as explained in Section 2.3, and then jumping between the models, as explained in this section. The entire process is iterated until convergence.

3. DATA ANALYSIS

3.1 ESTIMATING SENSITIVITY, SPECIFICITY, AND DISEASE PREVALENCE

The Appendix presents a data set comprising test results for 25 cows from an English farm. From a biological standpoint, mastitis is related not to the cow, but rather to the individual udder quarters, which are individual units. Thus for the purpose of this analysis, it is assumed that 25 cows result in 100 independent samples.

Table 4. Data matrix resulting from the thresholds 100 for the SCC test, 400 for the MAA test and 5.5 for the MEC test.

$n_{0000} = 41$	$n_{1000} = 7$	$n_{0100} = 1$	$n_{1100} = 1$
$n_{0010} = 2$	$n_{1010} = 2$	$n_{0110} = 2$	$n_{1110} = 3$
$n_{0001} = 26$	$n_{1001} = 8$	$n_{0101} = 0$	$n_{1101} = 0$
$n_{0011} = 0$	$n_{1011} = 0$	$n_{0111} = 1$	$n_{1111} = 6$

The bacteriological test differs from the other diagnostic tests in one main element. The bacteriological test is a qualitative test, which makes it easy to distinguish between a positive test result and a negative test result. A test is positive if a bacterium grows and negative otherwise. The other three tests are quantitative tests, requiring a threshold to distinguish between negative and positive test results. For the analysis of this data set, the usual thresholds were used: 100 for the SCC test, 400 for the MAA test, and 5.5 for the MEC test. These thresholds led to the $2 \times 2 \times 2 \times 2$ data matrix given in Table 4.

This data matrix served as the input in the algorithm described in the previous section to get estimates of the sensitivities and specificities of the four tests, as well as of the disease prevalence. Because no prior information about the model quantities was available, a Beta(1, 1) prior was used for the sensitivities and specificities of the four tests as well as for the prevalence. (See [Enøe, Georgiadis, and Johnson 2000](#) for a discussion on the specification of priors.) Three Markov chains each with 20,000 iterations but with different starting values were generated, and the tools described by [Toft et al. \(2007\)](#) and [Gilks, Richardson, and Spiegelhalter \(1995\)](#) were used to assess convergence of the chains. This included plotting the running mean of multiple sequences with overdispersed starting points and analyzing the Gelman–Rubin statistic and the autocorrelations. After an initial burn-in of 2000 iterations was discarded, only every seventh sample of each chain was saved because of high autocorrelations. These samples of the three chains were then combined to generate summary statistics.

The summary statistics of the combined chains are given in Table 5. The first test (with parameters Se_1 and Sp_1) is the bacteriological test, the second test (with parameters Se_2 and Sp_2) is the SCC test, the third test (with parameters Se_3 and Sp_3) is the MAA test, and the fourth test (with parameters Se_4 and Sp_4) is the MEC test. Note that in the summary statistics, the standard deviations of all four sensitivities are significantly greater than the standard deviations of the specificities. This indicates that the estimates for the specificities are more accurate than those for the sensitivities. This is discussed further in Section 4.

The summary statistics also reveal that the MEC test has poor accuracy compared with the other three tests. But here we are comparing the test accuracies only for a specific set of thresholds; an ROC analysis is needed to identify which test is the overall most accurate. This analysis is presented in Section 3.2.

Finally, we are interested in the dependence structure of the SCC and MAA tests. More can be learned about the dependence structure by analyzing the posterior densities shown in Figure 1. The posterior density of the parameter “model” indicates that the chains tend to remain longer in model 4 (the full conditional dependence model) than in the other three models. The model probabilities are 0.17 for model 1, 0.21 for model 2, 0.26 for model 3,

Table 5. Summary statistics of the combined chains.

	Mean	sd	2.5%	97.5%
π	0.17	0.06	0.07	0.32
Se ₁	0.76	0.13	0.49	0.98
Se ₂	0.73	0.17	0.36	0.98
Se ₃	0.78	0.16	0.40	0.99
Se ₄	0.51	0.14	0.25	0.79
Sp ₁	0.82	0.05	0.72	0.94
Sp ₂	0.96	0.03	0.89	1.00
Sp ₃	0.95	0.03	0.87	1.00
Sp ₄	0.61	0.05	0.50	0.71
$p_{\cdot 11 \cdot D}$	0.62	0.18	0.24	0.91
$p_{\cdot 00 \cdot \bar{D}}$	0.92	0.04	0.83	0.98
d_{Se}	0.32	0.35	0	0.96
d_{Sp}	0.34	0.34	0	0.95

NOTE: The sample means and standard deviations are given in the first two columns, and quantiles are given in the other columns. The medians of the disease prevalence, the sensitivities, the specificities, and the parameters $p_{\cdot 11 \cdot | D}$ and $p_{\cdot 00 \cdot | \bar{D}}$ lie within 0.03 of the mean. The medians of the parameters d_{Se} and d_{Sp} lie within 0.15 of the mean.

and 0.36 for model 4. Thus the SCC and MAA tests seem likely to have some degree of conditional dependence. In what follows, we evaluate whether or not this tendency is significant.

As suggested by Black and Craig (2002), the degree of dependence can be evaluated by analyzing d_{Se} and d_{Sp} , defined as follows:

$$d_{Se} := \frac{p_{\cdot 11 \cdot | D} - Se_2 Se_3}{\min(Se_2, Se_3) - Se_2 Se_3},$$

$$d_{Sp} := \frac{p_{\cdot 00 \cdot | \bar{D}} - Sp_2 Sp_3}{\min(Sp_2, Sp_3) - Sp_2 Sp_3}.$$

These parameters, which range between 0 and 1, represent the degree of dependence, with 0 indicating independence and 1 indicating complete dependence. These parameters were added to the MCMC simulation, leading to the summary statistics given in Table 5. We compared these distributions with the corresponding distributions resulting from independent data by comparing their means and medians. To do this, we generated 100 samples from the independence model with $\pi = 0.17$, $Se_1 = 0.76$, $Se_2 = 0.73$, $Se_3 = 0.78$, $Se_4 = 0.51$, $Sp_1 = 0.82$, $Sp_2 = 0.96$, $Sp_3 = 0.95$, and $Sp_4 = 0.61$ and computed the posterior densities of the parameters d_{Se} and d_{Sp} for each sample using the Markov chain simulation described earlier.

The values for d_{Se} and d_{Sp} obtained from the real data are quite likely under the independence model, because

$$p(\text{mean of } d_{Se} \text{ from independence model} \geq \text{mean of } d_{Se} \text{ from real data}) = 0.07,$$

$$p(\text{mean of } d_{Sp} \text{ from independence model} \geq \text{mean of } d_{Sp} \text{ from real data}) = 0.09,$$

$$p(\text{median of } d_{Se} \text{ from independence model} \geq \text{median of } d_{Se} \text{ from real data}) = 0.06,$$

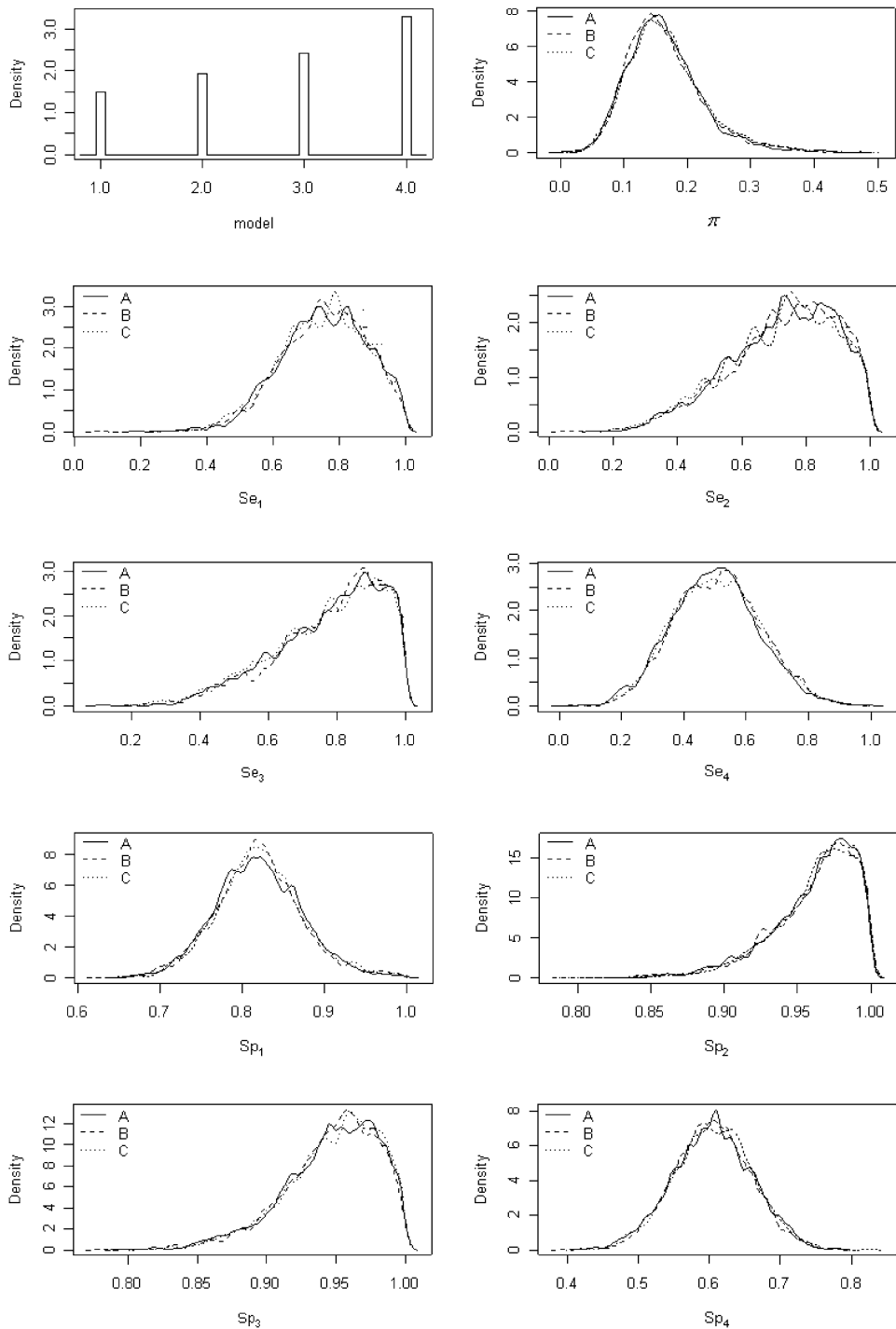


Figure 1. Posterior densities of each parameter for the three Markov chains A, B, and C.

and

$$p(\text{median of } d_{Sp} \text{ from independence model} \geq \text{median of } d_{Sp} \text{ from real data}) = 0.07.$$

Therefore, the assumption of independent tests seems reasonable. Because we assumed only a probable dependence between the SCC and MAA tests, we can conclude that all tests are independent.

3.2 ROC ANALYSIS

For the foregoing analysis, the decision thresholds were fixed (at 100 for the SCC test, 400 for the MAA test, and 5.5 for the MEC test). It is important to note that the sensitivity can be improved by lowering the value of the decision threshold, that is, by making the criterion for a positive test less strict. On the other hand, the specificity can be improved by increasing the value of the decision threshold, that is, by making the criterion for a positive test more strict. Thus the sensitivity and specificity of a test are inherently linked; as one increases, the other decreases. So when describing a diagnostic test, both sensitivity and specificity must be reported, along with the corresponding decision threshold. An ROC curve is a plot of the sensitivity against the *false-positive rate*, defined as *1-specificity*. Each point on the graph is generated by a different decision threshold. Thus the ROC curve describes the diagnostic accuracy of a test apart from the decision thresholds. (See [Zhou, Obuchowski, and McClish 2002](#) for an introduction to the ROC curve.)

The empirical ROC curves for the three quantitative mastitis diagnostic tests—the SCC test, the MAA test, and the MEC test—are presented in Figures 2, 3, and 4. The different decision thresholds are denoted by c . The points corresponding to the decision thresholds used for the summary statistics in Section 3.1 are printed in bold. Because the bacteriological test is a qualitative test, the decision threshold, and consequently the sensitivity and specificity, of the test are invariant; therefore, drawing the ROC curve makes no sense for this test.

The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The closer the curve comes to the diagonal of the ROC space, the less accurate the test. Therefore, the area under the ROC curve is a measure of test accuracy. It summarizes the accuracy of a test by a single number and is invariant to the prevalence of the disease. The area under the ROC curve can be computed very easily by constructing trapezoids under the curve and summing their areas. Table 6 gives these areas for each quantitative test. The table shows that accuracy of the SCC and MAA tests was excellent, whereas that of the MEC test was very poor.

Given the ROC curves in Figures 2, 3, and 4, computing the optimal decision threshold would be very interesting. Note that the optimal threshold depends on the disease prevalence and economic costs. It is remarkable that the standard decision thresholds of the SCC test ($c = 100$) and the MAA test ($c = 400$) result in operating points situated in the lower left of the ROC curve. These decision thresholds are optimal only if the prevalence is small and/or if treatment for the disease is harmful to healthy individuals and of little benefit to diseased individuals.

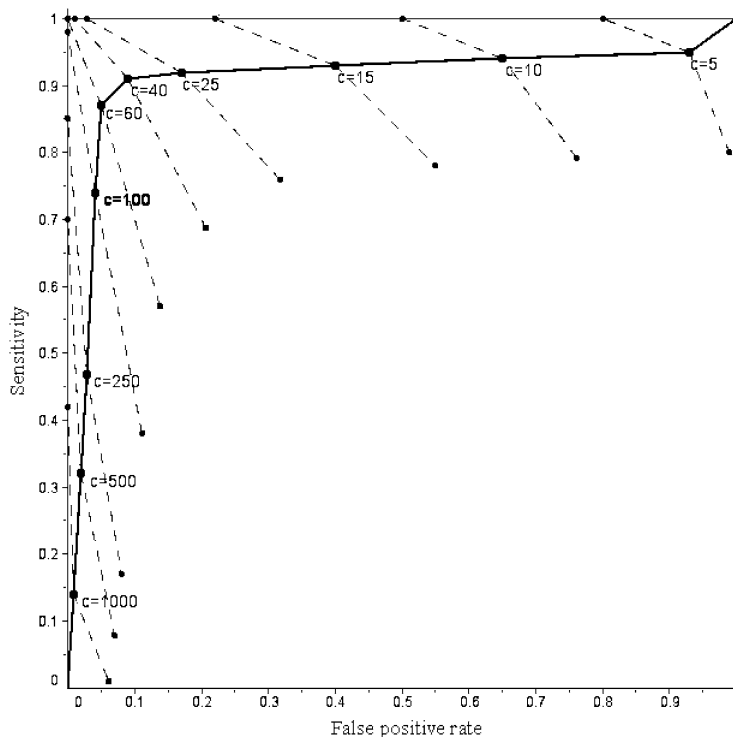


Figure 2. ROC curve of the SCC test with 95% credible interval.

4. DISCUSSION

The farm analyzed in this study had a quarter prevalence of mastitis of 17%. In other studies, the quarter prevalence of mastitis in Europe has been estimated to be about 20%. Thus the farm analyzed herein was close to the average.

The MEC test was cheap to perform but, as expected, exhibited very low sensitivity and specificity. The high probability of false negatives and false positives stems from the fact that electrical conductivity is influenced not only by mastitis, but also by many other factors, including the course of lactation, milking intervals, milk fat content and temperature, and foods ingested. The bacteriological test achieved only moderately good results and is very expensive. The MAA and SCC tests have the same moderate costs and demonstrated very good accuracy.

The summary statistics reveal a relatively high standard deviation of the sensitivities and show that the estimates of the sensitivities were less accurate than those of the specificities. This can be explained by the fact that information concerning sensitivity can come only from infected quarters, and because quarter prevalence was estimated as about 17%, only about 15–20 quarters were actually infected. This is too small a sample size to achieve good estimates with a small standard deviation. In future studies, this analysis should be repeated with a larger data set.

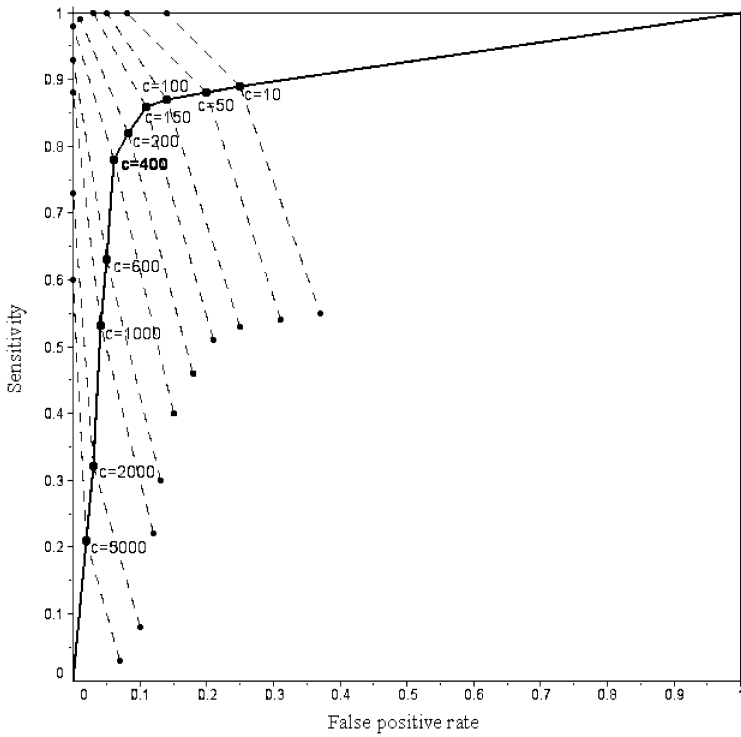


Figure 3. ROC curve of the MAA test with 95% credible interval.

The ROC analysis reveals minimal differences in the overall test accuracy of the SCC and MAA tests. But because of the relatively high standard deviation of the estimates and the fact that further data points and estimates could change the area under the ROC curve slightly in either direction, a definitive statement on the overall accuracy of these tests is not possible. The ROC analysis also shows that, depending on the decision threshold, the MAA or SCC test could perform better; however, the optimal decision threshold can be determined only with knowledge of the economic costs of treating a false-positive compared with not treating a false-negative. Unfortunately, such data remain unavailable.

The present analysis is based on the assumption that the four quarters of a cow are independent units and thus result in four independent samples. This assumption was verified using the chi-squared goodness-of-fit test. The resulting p -value was significant for the bacteriological test ($p = 0.012$) and for the MEC test ($p = 0.008$), whereas the null hypothesis of independent quarters was not rejected for the other two tests. As explained earlier, the electrical conductivity is affected not only by mastitis infection, but also by many other factors including the course of lactation, milking intervals, milk fat content and temperature, and food types ingested. These factors influence all four quarters similarly. This may explain the highly significant p -value for the MEC test. But the MEC test has very poor accuracy and thus is used only rarely to diagnose mastitis in practice. The bacteriological test also had a significant p -value, but taking into account the dependence of the quarters would result in 25 independent samples (1 for each cow), instead of 100

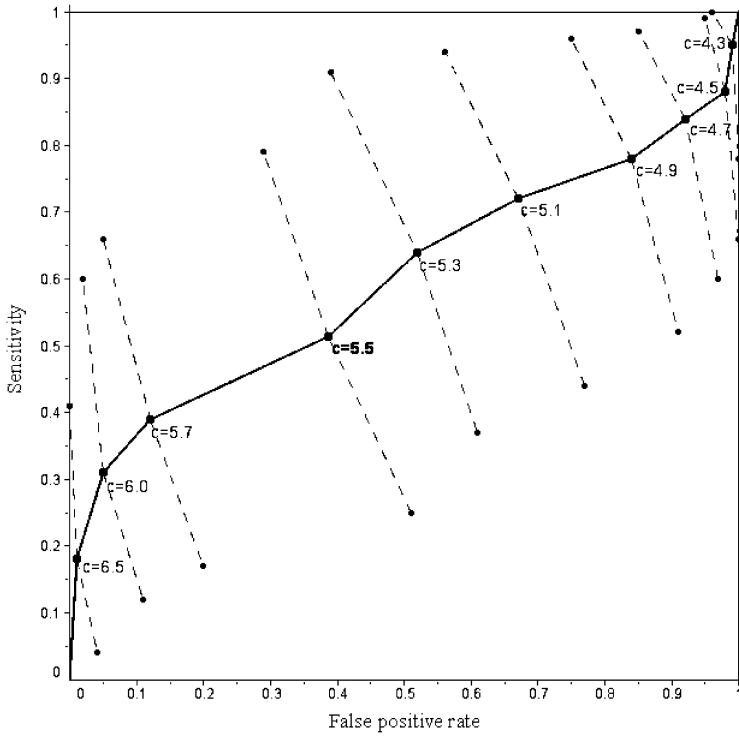


Figure 4. ROC curve of the MEC test with 95% credible interval.

samples. An analysis on the cow level instead of on the quarter level would require more data, which are not available. To overcome this problem, one could also formulate more complex models, such as a mixed-effects model with an additional parameter for the cow, gathering the correlation between the quarters.

When the results of four or more diagnostic tests are available, alternative methods of estimating the test accuracy are available, such as ordinary likelihood and the EM algorithm. We chose a Bayesian approach here because using RJMCMC not only provides estimates of the test accuracy and the disease prevalence, but also allows an analysis of the dependence structure between the SCC and MAA tests. Note that the ample degrees of freedom in our model compared with the original model of Black and Craig results from the fact that only partial pairwise dependencies are assumed and tested. Testing all pairwise dependencies would lead to a lack of degrees of freedom, similar to the model of Black and Craig. Thus our proposed model is useful when partial dependencies are suspect.

Table 6. Areas under the ROC curves for each quantitative test.

SCC test:	0.91
MAA test:	0.89
MEC test:	0.60

Further research requires the collection of more data, with the prime aim of achieving more precise estimates. In a first step, it is most important to collect data on the overall economic damage of mastitis. There is little scientific literature concerning the economics of mastitis, and much of what exists focuses on one particular treatment or element and its economic damage or benefit. Moreover, most studies use only simulation modeling, rather than actual data, to estimate the economic effects of mastitis. (See the discussion in [Hogeveen 2005](#), which reflects the current knowledge on mastitis from all over the world, as presented during the 4th IDF International Mastitis Conference in 2005.)

Farmers seem to have little awareness of the economic damage caused by mastitis, because these are mostly hidden costs. For example, the bulk of the damage is caused by diminishing milk production, and many farmers believe this effect to be linked to a cow's age rather than to mastitis. Collecting data on the economic damage of mastitis is a prerequisite for making farmers aware of the importance and the actual effects of mastitis in dairy production. Once aware, farmers will be better motivated to perform the various diagnostic tests for mastitis and to automatically supply data for a better-supported analysis of these tests. This would allow an analysis on the cow level instead of on the quarter level, resulting in a more precise estimation of the sensitivity and specificity of the various diagnostic tests, providing more reliable information on these tests' accuracy.

Finally, computing the costs of true/false positive/negative results would allow selection of the optimal decision threshold for each quantitative test and, consequently, a proper comparison of the different mastitis tests. Thus, collecting data on the losses is strongly recommended to estimate the economic effects and identify the overall optimal decision thresholds to help minimize the economic damage of mastitis in dairy production.

APPENDIX: DATA

This table gives the data set used for the analysis of mastitis in this study. The 25 cows from an English farm were tested at each quarter. The quarters are denoted by RF (right front quarter), RH (right hind quarter), LH (left hind quarter) and LF (left front quarter). Each cow was tested with the bacteriological test (Bact. test), the SCC test, the MAA test, and the MEC test. Because the bacteriological test is a qualitative test, a 0 indicates a negative test result. In most cases, mastitis is caused by bacteria, such as streptococci (strep), staphylococci (staph), coagulase-negative staphylococci (CNS), or corynebacteria (Cy). The number in front of the name of the bacteria denotes how many colonies grew. The other three tests are quantitative tests.

Cow no.	Quarter	Bact. test (qualitative test)	SCC (1000 cells/ml)	MAA (ng/ml)	MEC (milliSiemens)
2914	RF	20 CNS	13	0	4.6
2914	RH	0	230	1752	4.9
2914	LH	0	17	0	4.6
2914	LF	0	19	0	4.8
2904	RF	0	10	0	4.9
2904	RH	0	12	147	4.9
2904	LH	0	6	179	5
2904	LF	0	7	0	5
332	RF	0	13	0	4.6
332	RH	0	23	12	4.8
332	LH	0	12	0	4.7
332	LF	0	18	0	5
9390	RF	0	3	0	5.2
9390	RH	0	9	321	5.7
9390	LH	0	12	0	5.4
9390	LF	0	18	0	5.2
9393	RF	0	1229	2377	10.1
9393	RH	0	51	2470	5.1
9393	LH	0	91	5959	4.8
9393	LF	36 Cy	64	1313	4.8
415	RF	0	19	0	5.5
415	RH	0	47	0	5.6
415	LH	1 CNS	15	0	5.6
415	LF	10 Cy	9	0	5.4
9313	RF	10 Cy	41	37	5.2
9313	RH	0	16	0	5
9313	LH	0	45	0	5
9313	LF	10 CNS	45	0	5.2
127	RF	100 CNS	151	1274	4.3
127	RH	0	9	0	4.5
127	LH	0	844	679	4.5
127	LF	16 CNS	87	1370	4.9
9015	RF	0	47	0	4.7
9015	RH	0	27	0	5.1
9015	LH	0	131	339	5.1
9015	LF	30Strep	2925	6150	6.6
8128	RF	0	42	0	5.1
8128	RH	0	13	273	5.1
8128	LH	0	28	19	5.2
8128	LF	0	26	190	5.3
1049	RF	200 Staph	569	14652	6.2
1049	RH	0	10	0	5
1049	LH	0	8	0	4.8
1049	LF	0	9	0	4.8
8204	RF	0	9	86	5.2
8204	RH	0	16	28	5
8204	LH	0	15	0	4.6
8204	LF	0	12	0	4.5
9099	RF	0	4	0	5.4
9099	RH	0	5	0	5.3
9099	LH	0	7	0	5.2
9099	LF	0	7	0	5.4

(Continued.)

Cow no.	Quarter	Bact. test (qualitative test)	SCC (1000 cells/ml)	MAA (ng/ml)	MEC (milliSiemens)
324	RF	0	16	98	5.5
324	RH	0	60	0	5.6
324	LH	100 CNS	65	0	5.6
324	LF	0	7	0	5.6
174	RF	0	30	0	5.1
174	RH	100 CNS	34	0	5.6
174	LH	0	53	0	5.5
174	LF	8 Cy	38	347	5.8
381	RF	0	27	0	5.5
381	RH	0	20	0	5.1
381	LH	0	20	0	5
381	LF	0	16	50	5
9165	RF	0	12	110	5.4
9165	RH	0	12	0	5.5
9165	LH	0	8	0	5.5
9165	LF	0	13	0	5.5
2913	RF	200 CNS	494	1985	5.5
2913	RH	36 CNS	4	0	5.5
2913	LH	0	8	9	5.5
2913	LF	100 CNS	774	2064	5.6
892	RF	0	7	0	5.5
892	RH	0	10	0	5.5
892	LH	0	12	0	5.5
892	LF	0	12	0	5.5
7319	RF	200 CNS	136	172	5.1
7319	RH	48 CNS	54	0	5.2
7319	LH	0	6	0	4.9
7319	LF	0	3	0	4.9
6290	RF	37 CNS	36	0	6.1
6290	RH	28 CNS	118	415	6.1
6290	LH	0	6	0	5.9
6290	LF	0	7	0	6.1
375	RF	200 CNS	67	147	5.5
375	RH	66 CNS	48	0	5.8
375	LH	200 CNS	224	636	5.8
375	LF	0	9	0	5.5
7203	RF	0	6	0	5.8
7203	RH	0	5	0	5.5
7203	LH	0	9	0	5.4
7203	LF	0	14	8	5.5
8368	RF	0	29	196	6
8368	RH	0	33	0	6
8368	LH	0	12	50	5.9
8368	LF	0	15	99	5.6
118	RF	30 Cy	63	0	5.4
118	RH	45 Cy	216	1650	5.4
118	LH	28 Cy	254	570	5.4
118	LF	38 CNS	62	0	5.4

ACKNOWLEDGMENTS

The author thanks Andrew Barbour and Adrian Röllin for many helpful discussions, Paul Torgerson for support and briefings on problems caused by mastitis in dairy production, and Tridelta PLC and Elizabeth Berry of the Institute for Animal Health, Compton, UK, for providing the data.

[Received July 2007. Revised April 2008.]

REFERENCES

- Black, M. A., and Craig, B. A. (2002), "Estimating Disease Prevalence in the Absence of a Gold Standard," *Statistics in Medicine*, 21, 2653–2669.
- Enøe, C., Georgiadis, M. P., and Johnson, W. O. (2000), "Estimation of Sensitivity and Specificity of Diagnostic Tests and Disease Prevalence When the True Disease State Is Unknown," *Preventive Veterinary Medicine*, 45, 61–81.
- Gardner, I. A., Stryhn, H., Lind, P., and Collins, M. T. (2000), "Conditional Dependence Between Tests Affects the Diagnosis and Surveillance of Animal Diseases," *Preventive Veterinary Medicine*, 45, 107–122.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1995), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.
- Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. (1999), "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382–401.
- Hogeveen, H. (ed.) (2005), *Mastitis in Dairy Production*, The Netherlands: Wageningen Academic Publishers.
- Hui, S. L., and Walter, S. D. (1980), "Estimating the Error Rates of Diagnostic Tests," *Biometrics*, 36, 167–171.
- Toft, N., Innocent, G. T., Gettinby, G., and Reid, S. W. J. (2007), "Assessing the Convergence of Markov Chain Monte Carlo Methods: An Example From Evaluation of Diagnostic Tests in Absence of a Gold Standard," *Preventive Veterinary Medicine*, 79, 244–256.
- Toft, N., Jørgensen, E., and Højsgaard, S. (2005), "Diagnosing Diagnostic Tests: Evaluating the Assumptions Underlying the Estimation of Sensitivity and Specificity in the Absence of a Gold Standard," *Preventive Veterinary Medicine*, 68, 19–33.
- Walkenhorst, M. (2004), "Eine gute Prävention erhält die Eutergesundheit," *Die Grüne*, 4, 38–39.
- Whyte, D. S., Johnstone, P. T., Claycomb, R. W., and Mein, G. A. (2004), "On-Line Sensors for Earlier, More Reliable Mastitis Detection," in *British Mastitis Conference 2004*.
- Zhou, X., Obuchowski, N. A., and McClish, D. K. (2002), *Statistical Methods in Diagnostic Medicine*, New York: Wiley.