Continuous-Data Diagnostic Tests for Paratuberculosis as a Multistage Disease

N. Toft,¹ S. S. Nielsen,¹ and E. Jørgensen²

¹Department of Large Animal Sciences, The Royal Veterinary and Agricultural University, Grønnegårdsvej 8, DK-1870 Frederiksberg C, Denmark

²Biometry Research Unit, Danish Institute of Agricultural Sciences, P.O. Box 50, DK-8830, Tjele, Denmark

ABSTRACT

We devised a general method for interpretation of multistage diseases using continuous-data diagnostic tests. As an example, we used paratuberculosis as a multistage infection with 2 stages of infection as well as a noninfected state. Using data from a Danish research project, a fecal culture testing scheme was linked to an indirect ELISA and adjusted for covariates (parity, age at first calving, and days in milk). We used the log-transformed optical densities in a Bayesian network to obtain the probabilities for each of the 3 infection stages for a given optical density (adjusted for covariates). The strength of this approach was that the uncertainty associated with a test was imposed directly on the individual test result rather than aggregated into the population-based measures of test properties (i.e., sensitivity and specificity).

(Key words: paratuberculosis, ELISA, Bayesian network, test evaluation)

Abbreviation key: FC = fecal culture, FC_{high} = fecal culture high, FC_{low} = fecal culture low, FC_{neg} = fecal culture negative, Map = *Mycobacterium avium* ssp. *paratuberculosis*, OD = optical density, Pr = probability.

INTRODUCTION

Traditionally, the interpretation of a diagnostic test is a positive or a negative test result and the uncertainty associated with the test usually is measured by the sensitivity and specificity (i.e., the proportions of truly diseased and truly nondiseased identified correctly by the test). In this approach, however, limiting assumptions include 1) a dichotomous disease definition; 2) a threshold (cut-off) clearly dividing positive and negative test results; 3) all positive (and negative) test results are equally positive (or negative); and 4) the animals' test response is not affected by relevant covariates.

Often the test is carried out merely to determine whether a specific condition is present to initiate a suitable intervention. For this purpose, dichotomizing the disease definition and test result is adequate. It might be worthwhile, however, to improve this approach when the disease or condition and the associated threshold is ambiguous, thus allowing tests to be used in a wider range of settings.

Consider, as an example, paratuberculosis: a chronic, slowly developing infection in cattle and other ruminants caused by *Mycobacterium avium* ssp. *paratuberculosis* (Map; Chiodini et al., 1984). The chronicity of infection makes simple definitions of disease difficult. Furthermore, the sensitivity of diagnostic tests varies depending on the stage of the disease (Nielsen et al., 2002c). A frequently adopted intervention for infected cows is culling as opposed to doing nothing. However, some infected cows never develop "clinical" paratuberculosis with diarrhea and concomitant emaciation. Some managers only cull those cows that experience clinical disease, whereas other managers would like to detect and cull subclinical animals that might transmit Map to herd mates, are less productive, or both. Thus, a general framework to devise an optimal test-and-cull policy for paratuberculosis would benefit from a test that allows for multiple classifications of infection. Nielsen et al. (2002b) suggested 3 stages of Map infection: noninfected cows, infected cows with predominating cell-mediated immune responses, and infected cows with predominating humoral immune responses. The infected cows with predominating cell-mediated immune responses are assumed to have reduced antibody titers during the primary cell-mediated immune responses. During humoral immune responses, antibody titers are expected to be elevated. Hence, the 3 "infection groups" defined above may be assumed to correspond to 3 "immuno-groups": noninfected (having no antibodies); infected, with reduced antibody titers; and infected, with elevated antibody titers. Validating such immuno-groups requires repeated testing using fecal culture (FC), which is time consuming and expensive. Fecal culture generally takes 12 wk and sampling requires extra work compared with a milk-based indirect ELISA. However, given that a link be-

Received January 25, 2005.

Accepted July 15, 2005.

Corresponding author: Nils Toft; e-mail: nt@dina.kvl.dk.

tween the immuno-groups and the infection groups exists, antibody testing could provide a tool for inference about the amount of bacterial shedding, and hence, be used as decision support to livestock producers rather than using the more cumbersome and expensive FC method.

Other influences (such as parity and stage of lactation) on optical densities (OD) of the milk ELISA have been demonstrated previously (Nielsen et al., 2002c). Statistical models should include these covariates to improve the interpretation of the test result. Furthermore, such models should be easily adapted to repeated measures of test results on individual cows because this might be an element of a future testing regimen. Standard statistical methods such as mixed linear normal models (e.g., PROC MIXED, v. 8.2; SAS Inst., Inc., Cary, NC) are well established for inference from such models. Rather than OD given infection status, for diagnostic inference the conditional distribution of the infection status given the observed OD and relevant covariates of the cow is required. This becomes possible with comparatively little effort using probabilistic expert systems, such as Bayesian networks (Cowell et al., 1999).

The objective of the current study was to demonstrate how continuous-test data (and covariates) can be used for diagnostic testing when diseases are allowed to have multiple stages, as exemplified by an ELISA used for classifying lactating dairy cows into the 3 infection stages of paratuberculosis as defined above.

MATERIALS AND METHODS

Sampling Frame

For the current study, farms from a region in Southern Jutland, Denmark, where the Danish dairy industry initiated a project on infectious diseases in 1998 (Andersen et al., 2000) were used. The region was defined by 4 postal codes and contained approximately 260 dairy herds in 1998. Among these herds, 110 participated in the project. Specifically, results from 5 herds in which all cows had tested negative for paratuberculosis in 3 consecutive rounds of FC, based on samples collected 1 yr apart, and 8 herds in which several cows from each herd were shedding Map were used. These latter 8 herds were visited every third month between December 1999 and December 2002. Fecal samples were collected from all cows present at the time of the visit. Milk samples were collected 11 times/yr from each herd through the routine milk production scheme.

Test Methods and Classification Scheme

Fecal samples were cultured for 12 wk and classified as either negative or positive with varying degrees of

bacterial growth. Positive cultures were confirmed by PCR detecting IS900. The test was described in detail in Nielsen et al. (2004). Based on these classifications, the cows were trichotomized according to their Map infection status: 1) cows in 5 herds that never had any culture-positive cows were assumed free of paratuberculosis and all tested cows from these herds were classified as fecal-culture negative (**FC_{neg}**); 2) from the 8 herds with known problems, FC-positive cows with some negative cultures and whose positive cultures always had only few counts of bacteria (<10 cfu/g of feces) were defined as fecal culture low (**FC_{low}**); and 3) from the 8 herds with known problems, FC-positive cows with elevated bacterial counts (≥10 cfu/g of feces) or many repeated FC-positive tests (with any nonzero counts of bacteria) were defined as fecal culture high (**FC_{high}**). The latter group of cows can be perceived as those failing to control the Map infection.

Culture-negative cows from the 8 herds with known problems were designated “fecal-culture unidentified” and excluded from further analysis. The poor sensitivity of the FC prohibited a reasonable diagnosis of these cows. Based on this classification, 3983 results from 737 cows were included in the study.

The milk samples were tested in an ELISA for antibody based on *Mycobacterium paratuberculosis* strain 18 (an *M. avium* ssp. *avium* strain) from Allied Monitor (Fayette, MO). This test was described in detail previously (Nielsen et al., 2001, 2002c). The test results of the samples were classified based on their OD value from the ELISA reader. The OD was corrected for interplate variation by subtracting the OD of a negative control tested in each ELISA plate. Information on DIM, current parity, and the age at first calving of the tested cows was extracted from the Danish Cattle Database.

Statistical Models

Initial analyses showed that the variance of the OD values increased with increasing mean value. Thus, the corrected OD were log-transformed to stabilize their variance. Initially, a mixed linear normal model was formulated using random effects to take into account repeated measurements. Subsequent tests of model assumptions showed that variance homogeneity could not be assumed, and a more detailed variance model had to be formulated. The detailed model specification follows.

Paratuberculosis is a chronic infection, and infection is often assumed to occur during calfhoo. Thus, age was included in the model by using the proxies parity, stage of lactation (Nielsen et al., 2002a), and age at first calving. Because the association between log OD and age did not seem to be linear, however, DIM and age at first calving were coded as ordinal variables. Days in milk was di-

Figure 1. Bayesian network representation of the statistical model (equation 1): discrete nodes are represented as single-bordered ellipses, whereas the continuous Gaussian nodes are represented as double-bordered ellipses. The design effects include DIM, fecal-culture type (FC), parity (P), and age at first calving (AC). Estimates of these effects are represented in $D \times FC$, $FC \times P$, and $P \times AC$ for the interaction terms of the fixed effects, and herd effect (cow) and residual variation (res). The nodes Fix and Rand are summation nodes, and logOD is the log transformed optical density (OD) of the milk ELISA.

vided into 0 to 1, 2 to 11, 12 to 27, 28 to 40, and >40 wk after calving. Age at first calving was dichotomized into ≤ 28 and >28 mo. Parity was modeled as first, second, and third or greater parity. Although these variables were expected to influence OD for FC_{low} and FC_{high} cows, the same effect was not expected for FC_{neg} cows. Therefore, interaction between FC classification and the age covariates was expected. Furthermore, variance heterogeneity was expected between FC types, necessitating residual variances being allowed to vary. Variation among cows within herds was included as a random effect of cows nested within herds.

The initial model contained all main effects and their first-order interactions and assumed homogeneous variance. Model selection was performed by adding variance heterogeneity among subgroups when significant (tested by the likelihood-ratio test, $\alpha = 0.05$) and sequentially removing nonsignificant (using a 2-tailed test with $\alpha = 0.05$) fixed effects (by using SAS type III test using Satterthwaite's approximation for calculation of the degrees of freedom for the test). This procedure produced the following model for the log-transformed OD [log(OD)]:

$$\begin{aligned} \log(OD)_{himn} = & \mu + FC_i + P_{J_{himn}} + DIM_{K_{himn}} \\ & + AC_{L_{him}} + (FC \times P)_{iJ_{himn}} + (FC \times DIM)_{iK_{himn}} \\ & + (P \times AC)_{J_{himn}L_{him}} + [COW(FC \times HERD)]_{him} + \varepsilon_{himn} \end{aligned} \quad [1]$$

In this model, $\log(OD)_{himn}$ is the log-transformed corrected OD of the n th recording of the m th cow in herd h within the i th FC type. FC_i is the systematic effect of the i th FC type, $i = 1, 2, 3$; $P_{J_{himn}}$ is the systematic effect of parity, $J = 1, 2, 3$, with $J = 3$ indicating cows with

parity >2 ; $DIM_{K_{himn}}$ is the systematic effect of the K th DIM group, $K = 1, \dots, 5$; $AC_{L_{him}}$ is the systematic effect of the age at first calving, $L = 1, 2$ (subscripts L_{him} indicate that age at first calving is constant for a cow); $[COW(FC \times HERD)]_{him} \sim N(0, \sigma_c^2)$ is the random effect of cow within herd and FC type; and $\varepsilon_{himn} \sim N(0, \sigma_\varepsilon^2[\exp(\mathbf{U}\delta)]_{iJ_{himn}})$, i.e., residuals are identically Normal distributed within each combination of FC type and parity with σ_ε^2 as the common intercept term of the residual variance, $\mathbf{U}$ as the design matrix reflecting the combination of FC type and parity and δ as an 8-dimensional vector of estimated parameters.

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus, the expression $\sigma_\varepsilon^2[\exp(\mathbf{U}\delta)]$ gives a 9-dimensional vector: the first 3 elements give the variance for parity 1, 2, and ≥ 3 for FC_{neg} cows, the next 3 elements give the same variances for FC_{low} cows, and the last 3 give variances for FC_{high} cows. Hence, given estimates of σ_ε^2 and δ the variance for FC_{neg} and first-parity cows can be calculated as $\sigma_\varepsilon^2 \exp(\delta_1 + \delta_3 + \delta_5)$, the variance for FC_{low} , first parity as $\sigma_\varepsilon^2 \exp(\delta_2 + \delta_3 + \delta_7)$, etc.

Bayesian Network Model

Parameter estimates from the statistical model from equation 1 were used directly to form a Bayesian network (also known as a probabilistic network; Cowell et al., 1999). The qualitative part of this Bayesian network is shown in Figure 1. Bayesian networks may be constructed from expert knowledge concerning the domain. However, a network that corresponded closely to the statistical model with intermediate nodes to model the 2-way interactions was preferred. A special kind of Bayesian network, a Continuous Gaussian graph, where nodes are allowed to be either discrete (ellipses with solid borders) or continuous (Gaussian; ellipse with double lined border) nodes were used (Cowell et al., 1999). The arrows from one node (the parent) to another (the child) indicate a description of the conditional distribution of the values in the node given the values of its parents. For example, in Figure 1, DIM and FC are

Table 1. Estimates of the fixed effects of the model describing the log-transformed corrected optical density (OD), using fecal culture types (FC), parity (P), DIM, and age at first calving (AC). The interaction terms have absorbed the main effects to ease the transformation from the model to a Bayesian network.

Interaction	Variable	Level ¹	Estimate	SE	P value
DIM × FC	Weeks				
	0–1	FC _{neg}	–0.60	0.05	<0.0001
	2–11		–0.67	0.03	<0.0001
	12–27		–0.69	0.03	<0.0001
	28–40		–0.66	0.03	<0.0001
	>40	FC _{low}	–0.62	0.03	<0.0001
	0–1		–0.25	0.07	0.0003
	2–11		–0.31	0.04	<0.0001
	12–27		–0.25	0.04	<0.0001
	28–40	FC _{high}	–0.15	0.04	0.0005
	>40		–0.04	0.05	0.3818
	0–1		0.12	0.07	0.0801
	2–11		0.03	0.04	0.4914
	12–27		0.13	0.04	0.0033
	28–40		0.25	0.05	<0.0001
	>40		0.40	0.05	<0.0001
P × FC	Parity				
	1	FC _{neg}	–0.08	0.03	0.0114
	2		–0.00	0.03	0.7335
	≥3		0	—	—
	1	FC _{low}	–0.45	0.05	<0.0001
	2		–0.17	0.04	<0.0001
	≥3		0	—	—
	1	FC _{high}	–0.57	0.04	<0.0001
	2		–0.28	0.04	<0.0001
	≥3		0	—	—
P × AC	Parity				
	1	≤28 mo	–0.05	0.03	0.0659
	1	>28 mo	0	—	—
	2	≤28 mo	0.02	0.03	0.5488
	2	>28 mo	0	—	—
	≥3	≤28 mo	0.08	0.03	0.0146
	≥3	>28 mo	0	—	—

¹FC_{neg} = Fecal culture-negative cows; FC_{low} = fecal culture-positive cows with low bacterial count (< 10 cfu/g of feces); FC_{high} = fecal culture-positive cows with elevated bacterial count (≥ 10 cfu/g of feces) or many repeat FC-positive tests (with any non-zero counts of bacteria).

“parents” of the D × FC node. The discrete nodes were used for modeling the categorical design variables from the model (e.g., whether parity was at level 1, 2, or ≥3). The first level of continuous nodes was used to represent the parameter estimates and the further levels to represent the subsequent summation of the random and fixed effects. The intermediate summation might have been omitted but facilitates the modeling when repeated measurements on the same cow are modeled. Each continuous node has an associated table with the means and variances for the Normal distribution for each possible configuration of the parent nodes, thus utilizing the standard errors as well as the estimated parameters. For example, the line in the table describing the distribution of the continuous node (FC × P) with fecal type FC_{neg} and first parity would contain the estimate of the parameter (P × FC)_(1,FCneg) = –0.08 from the model in equation 1 as mean and the square of the corresponding SE (= 0.03) as variance (Table 1). The res node (representing the residual) would have a mean of 0 and the square of

the SD [= 0.26 for fecal type FC_{neg} and parity 1 (Table 2)] as variance. In Figure 1, the intercept and the main effects are pooled within the interactions.

In each node of a Bayesian network the observed value of a variable (evidence; i.e., the parity of a cow or the log OD value), may be entered. Using probabilistic calculus, the network reevaluates the probability distribution of the nodes in the network based on the evidence. This reevaluation is termed propagation of evidence. As an example, given evidence on every design variable, the network calculates the predicted log OD value with corresponding standard error. If no evidence is available, the network uses the prior distribution for the calculation. Prior distributions are often chosen to be noninformative (for example, equal probabilities for each level of the design variable). If evidence in the parity node were omitted, the calculated log OD would essentially correspond to least squares means in SAS terminology. As with least squares means, it is typically better to use a weighting that corresponds to the distribution over

parity in the population rather than equal weighting. In the Bayesian network, this corresponds to assigning a prior distribution to the parity node that corresponds to population distribution. When evidence (information) about a variable is entered into a Bayesian network, the prior distribution for that variable is disregarded. The strength of the Bayesian network is that it is possible to make deductions against the arrows. Upon entering evidence about the log OD, a revised probability distribution of the nodes without evidence is obtained. In the case of diagnostic information, the probability of a cow belonging to the FC_{neg} , FC_{low} , or FC_{high} category based on the measured log OD value is the primary concern. Thus, using a Bayesian network allows inference about the probability ($\Pr$)(FC_{neg}), $\Pr(FC_{low})$, and $\Pr(FC_{high})$ for a given log OD, using age at first calving, parity, and stage of lactation for a cow as age covariates, while accounting for the uncertainty associated with the parameter estimates. The inference made by propagations in the Bayesian network is essentially an application of Bayes formula (Cowell et al., 1999). Thus, it was possible to specify the “natural” model (i.e., FC type explaining log OD) in a traditional setting of mixed models ANOVA using a frequentist approach with well-established methods for model selection, and subsequently obtain the relevant information (log OD explaining FC type).

Because age at first calving, parity, and stage of lactation are known, evidence will overrule the prior, and uniform distributions may be used. The prior distribution on the proportion of FC_{neg} , FC_{low} , and FC_{high} cows should be given more attention, ideally reflecting the distribution in the population where the Bayesian network will be applied. Distribution in data, however, is not necessarily representative; a herd-specific prior might even be required. Hence, for the illustrative purposes of this study, the prior distribution of FC type is also assumed uniform.

RESULTS

In Figure 2, histograms for the log OD of each FC type and the empirical distribution of FC_{neg} , FC_{low} , and FC_{high} with increasing log OD are given. The probability that a cow with a log OD of -1 is FC_{neg} is nearly 40% [the vertical distance between the lower horizontal axis and the first (dotted) line], whereas $\Pr(FC_{low})$ for log OD above 0 is near 0%. The probability of FC_{low} is almost constant, but $\Pr(FC_{high})$ increases with increasing log OD. The empirical distribution is shown using stacked probability plot to emphasize that $\Pr(FC_{neg}) + \Pr(FC_{low}) + \Pr(FC_{high}) = 1$ for any given log OD.

Table 1 gives the parameter estimates and SE used in the quantitative part of the Bayesian network (Figure 1). The parameterization of the Bayesian network was

done without the main fixed effects. Hence, these were excluded (and consequently absorbed by the interaction terms) when estimating the solutions to the fixed effects. Although significant effects of the statistical model in equation 1 and the estimated parameters (Table 1) were only used for constructing the Bayesian network (Figure 1), it seems relevant to review briefly whether the results found here are in accordance with results found in previous studies. Using SAS Type III test with Satterthwaite's approximation to calculate the degrees of freedom for the tests, significant effects of the interaction between the DIM and fecal-culture type ($P < 0.001$), between fecal culture-type and parity ($P < 0.001$), and between parity and age at first calving ($P < 0.002$) were found. The mean log OD is increasing from FC_{neg} to FC_{low} and FC_{high} ; all 3 means were significant when tested using differences in least squares means (results not shown). This supports the hypothesis that the 3 classifications (FC_{neg} , FC_{low} , and FC_{high}) represent 3 different stages of the infection. In the presentation of the estimates, as well as in the Bayesian network, the possible correlation between the estimates has been ignored.

The estimate of the herd effect was $\sigma_C^2 = 0.0712$ (SE = 0.0053; $P < 0.0001$). Given the estimates of the common intercept term of the residual variance ($\sigma_\varepsilon^2 = 0.14$) and the vector $\delta = (-0.84, -0.32, 0.12, 0.30, 0.02, -0.96, 0.30, -0.18)$ the residual variance for the individual combinations of FC type and parity was calculated as:

$$\sigma_\varepsilon^2[\exp(U\delta)] = (0.0698, 0.0316, 0.0610, 0.1565, 0.1161, 0.1030, 0.1594, 0.1917, 0.1414).$$

The square roots of these estimates are given in Table 2. The estimated SD showed more variation in the log OD for FC_{high} than FC_{low} and FC_{neg} cows within each parity group. The main difference in variance, however, was due to differences between FC_{neg} and the other 2 classes (as expected). Variation between measurements of cows greater than second parity seemed to be less than between first-parity cows.

Using the estimates of Table 1, the random herd effect, and the calculated residual SD (Table 2), the quantitative part of the Bayesian network shown in Figure 1 was constructed. Using this network, the distribution of FC_{neg} , FC_{low} , and FC_{high} for any given log OD, parity, age at first calving, and DIM could be estimated. As each combination gives a different probability (and log OD is a continuous variable), the results of these propagations are best presented as graphs (Figure 3) using stacked plots to emphasize that $\Pr(FC_{neg}) + \Pr(FC_{low}) + \Pr(FC_{high}) = 1$ for any given combination of log OD and age covariates. In Figure 3, the probabilities of FC type for a given log OD are given for 9 different combinations

Figure 2. Histograms of the log-transformed corrected optical density (OD) of the ELISA by fecal culture (FC) types, and a plot of the empirical distribution of the 3 FC types as a function of the log-transformed corrected OD. $\Pr(\text{FC}_{\text{neg}})$ of a given $\log(\text{OD})$ is the distance between the lower horizontal axis and the dotted line, the distance between the dotted line and the dashed line is $\Pr(\text{FC}_{\text{low}})$, and the distance between the dashed line and the solid upper horizontal line is $\Pr(\text{FC}_{\text{high}})$.

Table 2. The estimated residual standard deviations for different combinations of fecal culture (FC) type and parity.

Parity	Fecal culture type		
	FC_{neg}	FC_{low}	FC_{high}
1	0.26	0.40	0.40
2	0.18	0.34	0.44
≥ 3	0.25	0.32	0.38

¹ FC_{neg} = Fecal culture-negative cows; FC_{low} = fecal culture-positive cows with low bacterial count (< 10 cfu/g of feces); FC_{high} = fecal culture-positive cows with elevated bacterial count (≥ 10 cfu/g of feces) or many repeat FC-positive tests (with any non-zero counts of bacteria).

of age covariates (the remaining combinations showed similar results, but were omitted to save space). For a given $\log \text{OD}$, $\Pr(\text{FC}_{\text{neg}})$, $\Pr(\text{FC}_{\text{low}})$, and $\Pr(\text{FC}_{\text{high}})$ differ substantially among parities. This is perhaps best seen when comparing small $\log \text{OD}$ and the associated probability $\Pr(\text{FC}_{\text{neg}})$ cow. Although the result for first-parity cows seemed inconclusive, the result for cows of third or greater parity indicated that small $\log \text{OD}$ are more likely to be the result of an FC_{neg} cow.

DISCUSSION

The statistical analysis confirmed the results from Nielsen et al. (2002b). The OD is affected by age, with an increase in $\log \text{OD}$ with age in general, but also within

Figure 3. Probabilities $Pr(FC_{neg})$, $Pr(FC_{low})$ and $Pr(FC_{high})$ as a function of the log-transformed corrected optical density (OD). Age at first calving is ≤ 28 mo for each plot, the vertical distance from the lower horizontal axis to the dotted line is $Pr[FC_{neg} | \log(OD)]$, the vertical distance between the dotted and the dashed lines is $Pr[FC_{low} | \log(OD)]$, and the vertical distance between the dashed line and the upper horizontal line is $Pr[FC_{high} | \log(OD)]$.

lactations. Allowing local effects of residual variances between FC types and parities confirmed that the variance of the ODs of noninfected cows (FC_{neg}) is smaller than for infected cows (FC_{low} and FC_{high}). The SD of the OD of FC_{neg} cows equals 40 to 80% of the SD of FC_{low} and FC_{high} cows when comparing FC types within parities.

Using a Bayesian network representation of the statistical model allowed the uncertainty associated with the estimates of the statistical model to be represented as variance on the Gaussian nodes. Hence, the estimates of probabilities provided by the Bayesian network are estimates based on the estimated parameters and their

associated uncertainty. This implies that the resulting probabilities gives a more realistic picture of the properties of the ELISA as a tool for classification of cows with respect to their bacterial shedding status than a traditional approach in which the misclassification is presented in terms of sensitivity and specificity of a test. Although the uncertainty associated with these estimates is often acknowledged, it is usually not used in the further interpretation of the results.

When using a trichotomized infection status rather than the usual dichotomization, the concepts of sensitivity and specificity lose their intuitive interpretation. Similarly, so does the use of likelihood ratios such as in Collins (2002), although, if the objective only was to handle continuous tests, they could be used. The Bayesian network and graphical representation of probabilities, however, could just as well be adapted when the traditional dichotomous disease classification is used (or generalized into more than 3 disease stages). Using Bayesian networks for decision support does not require construction of stacked probability plots, which are only used here for illustrative purposes. The statistical model showed that covariates influence the relationship between FC type and OD. Such covariates should be taken into account when interpreting the diagnostic test. This is further discussed in Nielsen and Toft (2002) for a traditional dichotomous disease definition. Consequently, computer-based methods are needed when interpreting the OD. This problem is not unique to the current study and rather than trying to further justify the use of covariates for test interpretation, it is probably more appropriate to challenge the justification in ignoring these known covariates. Visual comparison of the empirical distribution in Figure 2 and the individual distributions of Figure 3 seemingly shows an effect (gain) by including covariates. The plots in Figure 3 differ visually between different configurations of parity and DIM category. Whether these fluctuations are large enough to alter a culling decision based on an OD is another question that will be left for future studies.

The diagnostic properties of tests for Map are rather poor when used for just 2 stages of infection, and adding a third infection stage is unlikely to improve this. Throwing away information by dichotomizing or trichotomizing the test interpretation will further decrease the value of the test. To illustrate this, consider the situation shown in the lower left plot of Figure 3 and assume that 2 cut-offs were introduced so that $\log OD \leq -0.5$ was interpreted as FC_{neg}, $-0.5 < \log OD \leq 0.5$ as FC_{low}, and $0.5 < \log OD$ as FC_{high}. Using this approach, it would be possible to estimate parameters describing the probability of correctly classifying into each of the 3 categories. However, each log OD below -0.5 would have the same probability of being correctly (or wrongly) classified,

whereas using the log OD directly would give more confidence of a cow being FC_{neg} when the log OD equals -1.0 compared with -0.55 .

Eventually, it comes down to whether an interpretation of a cow being 70% FC_{neg}, 20% FC_{low}, and 10% FC_{high} is manageable. The alternative is to classify the cow as 1 of the FC types and accept that misclassification occurs. When used in a computer-based decision support system, there is no need to know the exact state of the cow with respect to infection with Map. Thus, there seems to be more benefit in representing the uncertainty associated with a given test result directly than through the traditional concept of test properties (such as sensitivity and specificity, which cannot be used here). Still, for a given set of conditions there is an optimal cutoff value for the OD in which all test results above that OD results in an action (such as culling). The point is that a continuous test (such as OD) combined with additional information allows this cutoff value to be assessed for different combinations of age at calving, parity, DIM, and with respect to a specific purpose of the test (i.e., control vs. eradication).

Testing and culling dairy cows according to their Map infection status should ideally be combined with the traits traditionally used to assess the value of a dairy cow in replacement models (e.g., milk yield, reproductive problems, and age); this is reviewed by Kristensen (1994). In Houben et al. (1994), a model that includes health traits such as mastitis were used for optimal replacement of dairy cows. Recently, Gröhn et al. (2003) developed a model that considered several different diseases when optimizing the replacement of dairy cows. These replacement models all assume that relevant information can be observed with certainty. This assumption, however, cannot be justified when addressing paratuberculosis. The true bacterial shedding state (or FC type) of the cow is generally not known and requires extensive and repeated testing (as described earlier in this paper) due to intermittent shedding and low diagnostic sensitivity in the early stages of infection. Fecal sampling and culturing is also more expensive and time consuming compared with milk sampling and testing using ELISA. Milk samples are already retrieved within the Danish milk-recording scheme for 88% of the Danish dairy herds. Thus, it might be tempting to apply the OD directly in the modeling without using the FC types. However, when modeling the effects of paratuberculosis on the risk of transmitting the infection to other animals, the FC types make more biological sense than the OD. Furthermore, methods for handling the partial observability within a traditional sequential decision-support framework with an infinite time horizon are being developed (Nilsson and Kristensen, 2002).

The present study uses a classification of cows into 3 different categories with respect to paratuberculosis

infection status. The procedure used for classification as well as the interpretation of the 3 groups can be questioned. The low sensitivity of the FC might introduce potential bias. However, the primary purpose of this study was to demonstrate that the use of dichotomized disease status to interpret a test result and the use of sensitivity and specificity to evaluate the assay method has shortcomings that make it worthwhile to consider other ways of interpreting and representing the uncertainty of diagnostic tests. The framework presented herein can easily be extended to allow for repeated tests, multiple diseases, or both. If the disease status cannot be determined by elaborate test schemes, then the statistical analysis needs to be carried out as a latent class analysis, such as the mixture model used in Nielsen et al. (2003).

ACKNOWLEDGMENTS

The Danish Dairy Board and the Kongeåproject are thanked for their financial support of the present study.

REFERENCES

- Andersen, H., K. Aagaard, F. Skjøth, E. Rattenborg, and C. Enevoldsen. 2000. Integration of research, development, health promotion, and milk quality assurance in the Danish dairy industry. Pages 258–260 in Proc. 9th Symp. Int. Soc. Vet. Epidemiol. Econ., Breckenridge, CO. Int. Soc. Vet. Epidemiol. Econ., Fort Collins, CO.
- Chiodini, R. J., H. J. van Kruiningen, and R. S. Merkal. 1984. Ruminant paratuberculosis (Johne's disease): The current status and future prospects. *Cornell Vet.* 74:218–262.
- Collins, M. T. 2002. Interpretation of a commercial bovine paratuberculosis enzyme-linked immunosorbent assay by using likelihood ratios. *Clin. Diagn. Lab. Immunol.* 9:1367–1371.
- Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. 1999. Probabilistic Networks and Expert Systems. (Statistics for Engineering and Information Science). Springer Verlag, New York, NY.
- Gröhn, Y. T., P. J. Rajala-Schultz, H. G. Allore, M. A. DeLorenzo, J. A. Hertl, and J. A. Galligan. 2003. Optimizing replacement of dairy cows: Modeling the effects of diseases. *Prev. Vet. Med.* 61:27–43.
- Houben, E. H. P., R. B. M. Huirne, A. A. Dijkhuizen, and A. R. Kristensen. 1994. Optimal replacement of mastitic cows determined by a hierarchic Markov process. *J. Dairy Sci.* 77:2975–2993.
- Kristensen, A. R. 1994. A survey of Markov decision programming techniques applied to the animal replacement problem. *Eur. Rev. Agric. Econ.* 21:73–93.
- Nielsen, S. S., C. Enevoldsen, and Y. T. Gröhn. 2002a. The *Mycobacterium avium* subsp. *paratuberculosis* ELISA response by parity and stage of lactation. *Prev. Vet. Med.* 54:1–10.
- Nielsen, S. S., Y. T. Gröhn, and C. Enevoldsen. 2002b. Variation of the milk antibody response to paratuberculosis in naturally infected dairy cows. *J. Dairy Sci.* 85:2795–2802.
- Nielsen, S. S., C. Grønbaek, J. F. Agger, and H. Houe. 2002c. Maximum likelihood estimation of sensitivity and specificity of ELISAs and faecal culture for diagnosis of paratuberculosis. *Prev. Vet. Med.* 53:191–204.
- Nielsen, S. S., H. Houe, S. M. Thamsborg, and V. Bitsch. 2001. Comparison of two ELISAs for serological diagnosis of paratuberculosis (Johne's disease) in cattle using different subspecies strains of *Mycobacterium avium*. *J. Vet. Diagn. Invest.* 13:164–166.
- Nielsen, S. S., B. Kolmos, and A. B. Christoffersen. 2004. Comparison of contamination and growth of *Mycobacterium avium* subsp. *paratuberculosis* on two different media. *J. Appl. Microbiol.* 96:149–153.
- Nielsen, S. S., and N. Toft. 2002. Optimisation of the validity of ELISA and faecal culture tests for paratuberculosis: Selection of population or correction by population characteristics? Pages 400–405 in Proc. 7th International Colloquium on Paratuberculosis, Bilbao, Spain. Int. Assoc. Paratuberculosis, Madison, WI.
- Nielsen, S. S., N. Toft, B. M. Bibby, and A. Nielsen. 2003. Estimation of seroprevalence of paratuberculosis with mixture models including covariates. Page 3 in Proc. Symp. Intern. Soc. Vet. Epidemiol. Econ., Vina del Mar, Chile. Int. Soc. Vet. Epidemiol. Econ., Fort Collins, CO.
- Nilsson, D., and A. R. Kristensen. 2002. Markov LIMID processes for representing and solving renewal problems. Pages 97–104 in Proc. 1st Eur. Workshop Sequential Decisions under Uncertainty in Agriculture and Natural Resources. Toulouse, France. INRA, Toulouse, France.