Genetic evaluation of susceptibility to clinical mastitis in Spanish Holstein cows

M. A. Pérez-Cabal,^{*1} G. de los Campos,[†] A. I. Vazquez,[‡] D. Gianola,^{†‡} G. J. M. Rosa,[‡] K. A. Weigel,[‡] and R. Alenda^{*}

^{*}Departamento de Producción Animal, Universidad Politécnica de Madrid, 28040 Madrid, Spain

[†]Department of Animal Sciences, and

[‡]Department of Dairy Sciences, University of Wisconsin, Madison 53706

ABSTRACT

This study had 3 objectives: to estimate genetic parameters and predict sires' transmitting abilities for clinical mastitis in a Spanish Holstein population, to propose a methodology for comparing models with different response variables by using a cost-based loss function, and to evaluate alternative genetic evaluation models by using this methodology. On-farm records for clinical mastitis from herds in 3 Spanish regions were analyzed as a binary trait (CM) and as number of episodes (NCM) per lactation. Linear and probit models were fitted for CM, whereas linear and Poisson models were used for NCM. Predictive ability of the models was evaluated by using the average predicted residual sum of squares from cross-validation and an alternative cost-based loss function. The loss function for model comparison was calculated by using average mastitis costs depending on the NCM and average cost per infected lactation. The average cost per infected lactation was \$345.58, whereas the cost per lactation ranged from \$204.86 to \$985.44 for lactations with 1 to 5 cases, respectively. Management and hygiene practices on individual farms had a large impact on clinical mastitis because the herd-year variance was larger than that of other random effects considered. The sire variance was significantly different from zero, confirming that genetic variation exists for clinical mastitis. Estimates of heritability for CM using the linear and probit models were 0.07 and 0.10 on the underlying scale, respectively. For NCM, the estimate of heritability for the linear model was 0.10 and estimates for the Poisson model evaluated at the mean and the median of lambda on the underlying scale were 0.09 and 0.07, respectively. Regarding ranking of sires, the definition of response variable (CM or NCM) was of greater importance than the choice of statistical model. Cross-validation results

indicated that models with the best fit for CM and NCM were the probit model and the linear model, respectively. However, a comparison across all models using the alternative cost-based loss function showed that using NCM as a response variable with a Poisson model provided the most accurate predictions of future costs associated with clinical mastitis.

Key words: clinical mastitis, genetic parameter, predicted cost, model comparison

INTRODUCTION

Mastitis is one of the most important diseases in the dairy industry because of its negative economic consequences and because of animal welfare and consumer concerns. Additional veterinary and labor costs, lower production, discarded milk, and involuntary culling reduce profitability (Shim et al., 2004; Wolfová et al., 2006). Strong worldwide selection pressure for increased milk production during the last 4 decades has led to high production levels, but also to deterioration in functional traits such as udder health and fertility. Therefore, genetic selection programs nowadays are increasingly focusing on reducing diseases and improving fertility traits (Rauw et al., 1998; Zwald et al., 2006).

Knowledge of genetic parameters of mastitis and assessment of the economic importance of the disease is required when developing a breeding program for improving udder health. Direct selection for resistance to clinical mastitis has been difficult, mainly because recording systems of diseases are not well developed in most countries. Whereas the Nordic countries have had well-established national health-recording systems since the 1970s and Norway was the first to publish EBV for clinical mastitis in 1978 (Carlén, 2008), most other countries are still struggling with data collection issues. Clinical mastitis is not routinely recorded on many commercial farms, and in herds where the trait is recorded, there is no mechanism for pooling data into a central database. Instead, indirect selection has been practiced widely, typically using SCC and conformation

Received December 16, 2008.

Accepted March 9, 2009.

¹Corresponding author: ma.perezcabal@upm.es

traits as indicators of resistance to mastitis (Shook, 1989; Hinrichs et al., 2005; Shook, 2006). In Spain, selection is based on these indirect traits by including them in the official selection index, ICO (Miglior et al., 2005), and a program was initiated in 2005 to begin on-farm collection of clinical mastitis records. Environmental effects and preliminary estimates of mastitis costs were reported in Perez-Cabal et al. (2008).

Clinical mastitis can be defined either as a binary (presence or absence, **CM**) or as a count variable (number of episodes, **NCM**). The binary response CM can be analyzed with linear or threshold models, whereas the linear, Poisson, and ordinal threshold models can be used for NCM. The most common approach has been the use of linear models assuming a normal distribution for CM. Resulting estimates of heritability have been low, ranging from 0.01 to 0.04 (e.g., Carlén et al., 2005; Interbull Center, 2008), partly because of a large influence of environment and management, but also because of small variation between cows for the binary trait. When nonlinear models were applied, heritability estimates for liability to mastitis were higher, ranging from 0.06 to 0.12 (e.g., Zwald et al., 2006; Vazquez, 2007; Negussie et al., 2008). A recent study by Vazquez et al. (2009) suggests that the definition of the trait (CM vs. NCM) is more important than the choice of model in terms of the ability to predict future performance. However, one difficulty is that models with different response variables cannot be compared using standard model comparison approaches.

The main aims of this study were 1) to estimate genetic parameters and predict transmitting abilities of sires for mastitis in a Spanish Holstein population; 2) to propose a method for comparing models with different response variables using a cost-based loss function, and 3) to evaluate alternative models in a Spanish Holstein population using that methodology.

MATERIALS AND METHODS

Data

Beginning in April 2005, clinical mastitis has been recorded in selected Spanish herds from the Basque Autonomous Region, Navarra, and Gerona. Clinical mastitis was diagnosed as secretion of visually abnormal milk from 1 or more quarters or inflammation of the udder tissue. For the same cow, 2 cases of mastitis were considered as different when the dates of diagnosis did not match and the treatment periods did not overlap, regardless of the number of quarters that were affected. Either farmers or their veterinarians recorded mastitis data by using specially designed data collection forms that were sent from regional milk-recording

associations to the Animal Production Department of the Polytechnic University of Madrid for data entry. The information recorded included herd, cow identification, quarter(s) affected, date(s) of the beginning (diagnosis) and end of treatment, treatment(s) applied, and remarks.

The data set used included records from 4,448 lactations of 3,261 cows that calved from April 2005 through January 2008 in 32 herds. These cows were daughters of 191 sires. Data editing requirements were a minimum of 5 daughters per sire, a minimum of 2 cows per herd-year of calving, and a minimum lactation length of 100 d.

Statistical Models

Clinical mastitis per lactation (for complete lactations, from calving to the end of lactation, and for incomplete lactations, from calving to 100 d after calving, at the least) was analyzed as CM, using a linear and a threshold (probit) mixed model, and as NCM (for 0, 1, 2, 3, 4, and 5 episodes per lactation; more than 5 cases per lactation were not considered), using linear and Poisson mixed models. The following linear predictor was common to all models:

$$\eta_{ijklm} = \alpha_0 + \alpha_1 A_{jm} + S_i + L_j + hy_k + sire_l + pe_m,$$

where η_{ijklm} is a linear predictor of susceptibility to clinical mastitis; α_0 is an effect common to all observations; A_{jm} is the age at calving of the j th lactation of the m th cow, and α_1 is a regression coefficient; S_i is the fixed effect of the i th calving season (4 levels: January to March; April to June; July to September; October to December); L_j is the fixed effect of the j th lactation (7 levels); hy_k is the random effect of the k th herd-year of calving (92 levels); $sire_l$ is the transmitting ability of the l th sire (191 sires); and pe_m is the permanent environmental effect of the m th cow (3,261 cows). In this study, mastitis in different lactations was then analyzed with a repeatability model. Alternatively, with larger data sets, they can be analyzed as different traits (Heringstad et al., 2004; Zwald et al., 2006).

The joint distribution of random effects included in the linear predictor was as follows:

$$\begin{pmatrix} \mathbf{hy} \\ \mathbf{sire} \\ \mathbf{pe} \end{pmatrix} \sim N \left[\mathbf{0}, \begin{pmatrix} \mathbf{I}\sigma_{hy}^2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}\sigma_{sire}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_{pe}^2 \end{pmatrix} \right],$$

where **hy**, **sire**, and **pe** are the vectors of herd-year of calving, sire, and permanent environmental effects, respectively, and σ_{hy}^2 , σ_{sire}^2 , and σ_{pe}^2 are the corresponding

variances; $\mathbf{A}$ is the additive relationship matrix between sires (order 1,158 by 1,158), built with information from a 7-generation pedigree, and $\mathbf{I}$ is an identity matrix of appropriate order.

The specification for the linear models for CM and NCM is completed as follows:

$$\begin{aligned} CM_{ijklmn} &= \eta_{ijklmn} + \varepsilon_{ijklmn}, \text{ and} \\ NCM_{ijklmn} &= \eta_{ijklmn} + \tilde{\varepsilon}_{ijklmn}, \end{aligned}$$

with the joint distribution of model residuals being $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma_{\varepsilon}^2)$ and $\tilde{\boldsymbol{\varepsilon}} \sim N(\mathbf{0}, \mathbf{I}\sigma_{\tilde{\varepsilon}}^2)$, respectively. The residual variances can be interpreted as reflecting variation between temporary environmental effects.

A threshold model using the probit link (hereafter denoted as the probit model) was also fitted to CM data, setting the threshold to 0. The scoring rule in this model was

$$CM_{ijklmn} = \begin{cases} 1(\text{presence}) & \text{if } \mu_{ijklmn} > 0 \\ 0(\text{absence}) & \text{otherwise} \end{cases},$$

where $\mu_{ijklmn} = \eta_{ijklmn} + z_{ijklmn}$ is a latent variable (liability to mastitis), and the z_{ijklmn} are independent and identically distributed standard normal random variables. In this model, the probability of an event of clinical mastitis is $P(CM = 1 | \mu_{ijklmn}) = \Phi(\mu_{ijklmn})$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

A Poisson model was fitted to NCM. With the log-link $\left(\lambda_{ijklmn} = e^{\eta_{ijklmn}}\right)$, the conditional probability of a given number of events is

$$P(NCM = t | \lambda_{ijklmn}) = \frac{e^{-\lambda_{ijklmn}} (\lambda_{ijklmn})^t}{t!},$$

for $t = 0, 1, 2, 3, 4, 5$.

Analyses were performed in R (R Development Core Team, 2008) using the *pedigreemm* package (Bates and Vazquez, 2008) to accommodate additive relationships among sires. Variance components were estimated using REML. In the linear models, the solution has a close form, but for the generalized linear models, the solutions were estimated using Laplace approximation to the marginal likelihood.

Heritability for the linear model was assessed as

$$h^2 = \frac{4\sigma_{sire}^2}{\sigma_{hy}^2 + \sigma_{sire}^2 + \sigma_{pe}^2 + \sigma_{res}^2},$$

where σ_{res}^2 was σ_{ε}^2 for CM and $\sigma_{\tilde{\varepsilon}}^2$ for NCM. Heritability for the probit model was

$$h^2 = \frac{4\sigma_{sire}^2}{1 + \sigma_{hy}^2 + \sigma_{sire}^2 + \sigma_{pe}^2}.$$

As mentioned in Vazquez (2007), estimates of heritability from the Poisson model depend on the value of λ . Therefore, heritability was evaluated at both the mean and the median fitted values of λ , following the approach of Foulley et al. (1987), using the expression

$$h^2 = \frac{4\sigma_{sire}^2}{\sigma_{hy}^2 + \sigma_{sire}^2 + \sigma_{pe}^2 + \lambda^{-1}}.$$

Permutation Test for Variance Components

All models included the random effects of grouping factors (sire, herd-year, and permanent environment) that, a priori, were thought to explain variability in susceptibility to clinical mastitis. The need to include these grouping factors in the model can be evaluated through hypothesis testing. The null hypothesis was that the corresponding variance parameter was zero. A standard likelihood ratio test cannot be used because the null hypothesis is at the edge of the parameter space. Therefore, a permutation test was performed, as in Fitzmaurice et al. (2007). For each of the 3 variance components, the distribution under the null hypothesis was estimated using 200 permutations obtained by randomly shuffling the identification of the corresponding grouping factor while keeping the other aspects of data structure unchanged. *P*-values were estimated as the relative frequency of permutations for which the estimated variance component was greater than or equal to the estimated variance in the original data set.

Evaluation of Predictive Ability via Cross-Validation

The ability of the models to predict independent observations was assessed using a 10-fold cross-validation. The assignment of observations into folds was at random. A detailed discussion regarding cross-validation is provided by Efron and Tibshirani (1993). The average predicted residual sum of squares (MPRSS) was computed as

$$MPRSS = \frac{1}{N} \sum_{f=1}^{10} \sum_{(ijklmn) \in f} \left[y_{ijklmn} - E(y_{ijklmn} | \hat{\eta}_{ijklmn}^{-f}) \right]^2,$$

where y_{ijklmn} is an observation for CM or NCM; $\hat{\eta}_{ijklmn}^{-f}$ is the estimated value of the linear predictor obtained by fitting the corresponding model to a data set that did not include the observations from the f th fold; $E(\cdot)$ is the expected value of y_{ijklmn} expected from the model with fixed and random effects excluding the observations from the f th fold; $(ijklmn) \in f$ indicates that the corresponding observation belongs to the f th fold; and N is the number of observations in the original data set. For model comparison, smaller values of MPRSS are better.

Although MPRSS can be used to compare models with the same response variable, it cannot be used to compare models with different outcome responses when the measure is based on different metrics. To compare across models with different response variables, a cost-based loss function was developed, as described below.

Cost-Based Loss Function

One way of comparing the predictive ability of different models is to evaluate how well they predict the costs associated with mastitis. A cost function that assigns monetary losses to different values of NCM was estimated from the actual mastitis cost of each cow in the studied population. Mastitis costs corresponding to each lactation with at least 1 infection in the data set were computed by considering drugs, discarded milk, reduced milk production, veterinary fees, and extra labor. A detailed description of how these costs were computed is given in the Appendix. The average cost by number of cases of mastitis, hereafter denoted as $Cost(NCM = t)$, where $t = 0, \dots, 5$, was estimated by averaging the estimated cost within values of NCM; by definition, $Cost(NCM = 0) = 0$. The cost of having at least 1 case of mastitis, $Cost(NCM > 0)$, was computed by averaging mastitis costs of cows that had at least 1 episode.

In the probit and Poisson models, the loss function was the absolute value of the difference between the actual and predicted costs, based on the observed number of events, and the probabilities of becoming infected or having t cases, respectively. In the linear models, the expected cost was computed by interpolation. Specifically, the loss function associated with the $ijklmn$ record for each model was computed as follows:

Linear model for CM:

$$loss_{ijklmn} = \left| CM_{ijklmn} - \hat{\eta}_{ijklmn} \right| Cost(NCM > 0),$$

Probit model for CM:

$$loss_{ijklmn} = \left| CM_{ijklmn} - \Phi(\hat{\eta}_{ijklmn}) \right| Cost(NCM > 0),$$

Linear model for NCM:

$$\left(loss_{ijklmn} \mid t \leq \hat{\eta}_{ijklmn} \leq t+1 \right) = \left| Cost(NCM_{ijklmn} = t) - \frac{Cost(t) + (\hat{\eta}_{ijklmn} - t) [Cost(t+1) - Cost(t)]}{1} \right|,$$

Poisson model for NCM:

$$\left(loss_{ijklmn} \mid NCM_{ijklmn} = t \right) = \left| Cost(NCM_{ijklmn} = t) - \sum_{t=0}^5 \frac{\hat{\lambda}_{ijklmn}^t e^{-\hat{\lambda}_{ijklmn}}}{t!} Cost(NCM_{ijklmn} = t) \right|,$$

where $CM_{ijklmn} = \{1 \text{ if presence; } 0 \text{ otherwise}\}$; $\hat{\eta}_{ijklmn}$ is the fitted value of the linear predictor for the corresponding record and model; $\Phi(\hat{\eta}_{ijklmn})$ is the estimated probability of having at least one episode of mastitis with the probit model for CM; and $\hat{\lambda}_{ijklmn}^t e^{-\hat{\lambda}_{ijklmn}} / t!$ is the fitted probability of having t cases of mastitis. Models were compared based on the mean value of the loss function.

RESULTS AND DISCUSSION

Descriptive Statistics

Average milk production was 10,441 kg in 332 d (Table 1). Actual milk yield in healthy and diseased lactations was 10,354 kg in 330 d, and 10,669 kg in 337 d, respectively. The higher milk production in the sick cows was probably due to the larger percentage of infected animals among the older cows (from the second lactation on). The average lactation number for joint groups was 2.19, whereas the average for infected and for healthy lactations was 2.57 and 2.04, respectively, with higher mastitis frequency in later lactations. The average NCM per lactation was 0.46 with a standard deviation of 0.91, and the recurrence of clinical mastitis, calculated as the ratio between number of cases and number of cows at risk, was 1.67. With regard to the distribution of NCM, 72% of the lactations were healthy and 28% of lactations were diagnosed with at least 1 case of clinical mastitis. Among the diseased lactations, the percentage of lactations with 1, 2, 3, 4, or 5 cases, was 59, 24, 10, 4, and 3%, respectively. The average number of daughters per sire was 17 (13 healthy cows and 4 infected cows), ranging from 5 to 145, with 74% of the sires having from 5 to 15 daughters and 4% of the sires with more than 80 daughters.

Table 1. Distribution over lactations and means (SD) for number of daughters per sire; DIM; actual milk, fat, and protein yield per lactation; clinical mastitis (CM); and number of cases of clinical mastitis (NCM) in all, healthy, and infected lactations

Item	All	Healthy	Infected
Records, n	4,448	3,223	1,225
Lactation 1, %	37.3	30.8	6.5
Lactation 2, %	29.7	21.2	8.5
Lactation 3, %	18.3	11.5	6.8
Lactation 4, %	8.7	5.5	3.2
Lactation 5, %	4.0	2.4	1.6
Lactation 6, %	1.5	0.7	0.7
Lactation 7, %	0.5	0.3	0.3
Average lactation number	2.19 (1.26)	2.04 (1.20)	2.57 (1.34)
DIM	332 (94)	330 (93)	337 (96)
Milk, kg	10,441 (3,283)	10,354 (3,254)	10,669 (3,352)
Fat, kg	373 (124)	371 (123)	381 (124)
Protein, kg	333 (107)	331 (106)	340 (110)
CM, %	28	—	—
NCM	0.46 (0.91)	—	1.67 (1.0)
Daughters per sire, n	17.07 (22.75)	13.00 (17.64)	4.07 (5.89)

Estimates of Genetic Parameters

Table 2 shows estimated variance components and heritability for each model, as well as results of the permutation tests. Herd-year, sire, and permanent environmental effects were significantly different from zero in all models, but the permanent environmental variance in the linear model for CM was not. This was not surprising, because there were few repeated lactations per cow, and the binary response variable does not capture variation well in the number of episodes between lactations of the same cow. For all models, the herd-year variance was the largest, highlighting the importance of management and hygiene practices on individual farms (e.g., Rodrigues-Motta et al., 2007; Vazquez et al., 2009). The estimated permanent environmental variance was larger than the variance between sires in the linear and Poisson models for NCM. The sire variance was significantly different from zero

($P < 0.0001$), confirming that genetic variation exists for clinical mastitis.

Estimated heritability for CM with the linear model was 0.07. Similar results were reported in Vazquez et al. (2009) for Norwegian Red cattle, although most studies have reported values ranging from 0.01 to 0.04 on the observed scale (Hansen et al., 2002; Carlén et al., 2005; Negussie et al., 2008). With the probit model, estimated heritability of CM on the underlying scale was 0.10. From the literature, threshold models for CM have typically produced higher estimates of heritability of liability to clinical mastitis than linear models, with estimates ranging from 0.06 to 0.12 (Heringstad et al., 2003; Zwald et al., 2006; Negussie et al., 2008). When mastitis was analyzed as NCM, the linear model estimate of heritability was 0.10. The Poisson model evaluated at the mean and median of lambda (0.4561 and 0.3069, respectively) led to estimates on the underlying scale of 0.09 and 0.07, respectively. These estimates

Table 2. Estimated variances (SE in parentheses) of herd-year (σ_{hy}^2), sire (σ_{sire}^2), permanent environmental (σ_{pe}^2), and residual (σ_{res}^2) effects, estimated heritability¹ (h^2), and P -values from permutation tests for the presence or absence of clinical mastitis (CM) and number of cases of clinical mastitis (NCM)

Model	σ_{hy}^2	σ_{sire}^2	σ_{pe}^2	σ_{res}^2	h^2
Linear (CM)	0.022** (0.000)	0.003** (0.000)	0.000 (0.002)	0.170 (0.002)	0.07 (0.007)
Probit (CM)	0.320** (0.003)	0.033** (0.003)	0.019** (0.003)	1.000 (—)	0.10 (0.013)
Linear (NCM)	0.117** (0.001)	0.021** (0.002)	0.065* (0.012)	0.624 (0.001)	0.10 (0.007)
Poisson (NCM)	0.714** (0.001)	0.079** (0.008)	0.567** (0.024)	—	0.09 ² /0.07 ³ (0.012/0.008)

¹Linear model: $h^2 = 4\sigma_{sire}^2 / (\sigma_{hy}^2 + \sigma_{sire}^2 + \sigma_{pe}^2 + \sigma_{res}^2)$; Probit model: $h^2 = 4\sigma_{sire}^2 / (1 + \sigma_{hy}^2 + \sigma_{sire}^2 + \sigma_{pe}^2)$; Poisson model: $h^2 = 4\sigma_{sire}^2 / (\sigma_{hy}^2 + \sigma_{sire}^2 + \sigma_{pe}^2 + \lambda^{-1})$.

²Heritability evaluated at mean $\lambda = 0.4561$.

³Heritability evaluated at median $\lambda = 0.3069$.

* $P < 0.01$; ** $P < 0.0001$.

Table 3. Estimated mastitis cost per lactation by number of cases of mastitis (NCM) or the presence or absence of clinical mastitis (CM), including the total cost, discarded milk cost, drug cost, extra labor cost, veterinary cost, and reduced milk production per lactation

Item	NCM					NCM ≥ 1 (CM = 1)
	1	2	3	4	5	
Total cost, \$	204.86	422.79	608.07	745.38	985.44	345.58
Discarded milk, \$	116.32	225.16	336.90	412.06	559.15	202.13
Drugs, \$	31.30	63.88	88.43	110.36	130.66	51.46
Labor, \$	7.12	18.69	22.54	27.82	35.46	13.25
Veterinary costs, \$	2.01	3.92	5.85	7.71	10.00	3.35
Reduced milk, \$	45.00	109.35	150.73	177.01	217.26	82.31
Reduced milk, kg	66	162	230	273	376	221

were similar to those reported by Vazquez et al. (2009) with linear and Poisson models.

Predicted Transmitting Abilities

Predicted transmitting abilities for CM ranged from -0.0785 to 0.0965 (SE ranged from 0.0266 to 0.0562) for the linear model, and from -0.2600 to 0.2494 (SE ranged from 0.0887 to 0.1804) in the liability scale. For NCM, the ranges were from -0.2184 to 0.3884 (SE from 0.2230 to 0.2443) for the linear model, and from -0.3924 to 0.4498 (SE from 0.1229 to 0.2779) for the Poisson model [in the $\log(\lambda)$ scale]. The lowest values of SE for all models were for sires with more than 80 daughters, so the accuracy of corresponding PTA was higher than for other sires with less information. As expected, the linear and probit models provided similar PTA for CM (correlation of 0.99). For NCM, PTA from the linear and Poisson models were highly correlated (0.92) as well, although predictions differed for sires with extreme PTA values. If selection programs addressed variation in the number of cases of mastitis per lactation, where the main interest is in discarding animals that produce offspring with frequent health problems, the sires selected using linear and Poisson models for NCM would differ depending on selection intensity. For example, for an intensity of 5% (top 10 sires), only 6 of 191 sires with daughters were in common between these models, whereas 18 sires were in common at an intensity of 10% (top 20 sires). However, Spearman correlations were smaller between linear models for CM and NCM (0.86) than between probit and Poisson models (0.91) for CM and NCM. This confirms that the definition of the response variable had a greater impact on sire rankings than the choice of model, which is in agreement with Vazquez et al. (2009).

Mastitis Costs

Mastitis costs are shown in Table 3. Total costs per lactation increased from \$204.86 for 1 case of mastitis

[$Cost(NCM = 1)$] to \$985.44 for 5 cases of mastitis [$Cost(NCM = 5)$], with an average of \$345.58 per infected lactation [$Cost(NCM > 0)$]. In this population, with an incidence rate of 28%, mastitis costs per 100 cows were estimated as \$9,675 per lactation. The most important source of costs was discarded milk, which represented 56 to 60% of total costs, followed by reduced milk yield (20 to 23%), drugs (13 to 15%), extra labor (3 to 4%), and veterinary fees (1%). Other studies reported average costs ranging from \$71 to \$215 per lactation, depending on country, severity level, age of cow, and estimation procedure (i.e., Shim et al., 2004; Wolfová et al., 2006). There is a consensus among studies that reduced milk yield and discarded milk are the most important components of cost (Bar et al., 2008; Guard, 2008). The loss in milk production of diseased cows as compared with healthy cows (Table 3) was 66, 162, 230, 273, or 376 kg for a cow with 1, 2, 3, 4, or 5 cases, respectively. The average loss in milk production across all infected lactations was 221 kg. Estimated yield losses in the literature range from 0 to 915 kg per lactation, depending on the parity and stage of lactation at the time of infection (Hortet and Seegers, 1998; Bar et al., 2007).

Model Comparison

Mean predicted residual sums of squares obtained in the cross-validation are shown in Table 4. The probit model for CM and the linear model for NCM provided the best fit for each corresponding trait. Healthy lactations (CM = 0 and NCM = 0) were predicted more accurately using nonlinear models (i.e., probit or Poisson). For CM, the predictive ability of linear and nonlinear models was similar, although the probit model led to slightly smaller MPRSS for healthy lactations and the linear model provided slightly smaller MPRSS for infected lactations. Parallel results were found for NCM, where MPRSS for healthy lactations was smaller for the Poisson model, whereas MPRSS for lactations with at least 1 case of clinical mastitis was smaller with

Table 4. Mean predicted residual sum of squares of cross-validation for the presence or absence of clinical mastitis (CM) and number of cases of mastitis (NCM) for the linear, probit, and Poisson models for all lactations, healthy lactations, infected lactations, and lactations with 1 to 5 cases of mastitis

Item	CM		NCM	
	Linear	Probit	Linear	Poisson
All	0.18	0.18	0.73	0.75
Healthy	0.09	0.08	0.27	0.23
Infected	0.42	0.42	1.95	2.12
NCM = 1	—	—	0.30	0.42
NCM = 2	—	—	1.84	2.04
NCM = 3	—	—	4.82	5.13
NCM = 4	—	—	9.40	9.39
NCM = 5	—	—	16.61	17.22

the linear model. Estimates of MPRSS using NCM as the response variable were computed as in Vazquez et al. (2009), and the results (not shown) followed the pattern described above.

Table 5 shows the estimated values of the loss function for the 4 models. Results indicated that using NCM as the response variable improved the ability of the model to predict future costs. The Poisson model for NCM predicted mastitis costs 21, 20, and 12% more accurately than the linear model for CM, the probit model for CM, and the linear model for NCM, respectively. Within models for the same response variable, the nonlinear models provided predictions that were more accurate.

CONCLUSIONS

Four alternative linear and nonlinear models for genetic evaluation of mastitis in a Spanish Holstein population were considered. Estimated genetic parameters were in agreement with literature values from other populations, and clinical mastitis had significant genetic variation, so selection for greater resistance to mastitis in the Spanish population should be effective.

Table 5. Loss function (in \$/lactation) for prediction of mastitis costs using presence/absence of clinical mastitis (CM) or number of cases of mastitis (NCM) as the response variable

Item	CM		NCM	
	Linear	Probit	Linear	Poisson
All	118.01	113.50	103.01	85.00
Healthy	83.35	78.43	71.47	55.37
Infected	212.69	209.29	189.15	165.94
NCM = 1	—	—	87.65	92.99
NCM = 2	—	—	244.81	211.07
NCM = 3	—	—	375.81	291.49
NCM = 4	—	—	469.58	342.84
NCM = 5	—	—	673.64	541.03

A new method for model comparison using mastitis costs was proposed, and results indicate that the definition of the trait used to evaluate clinical mastitis (CM vs. NCM) should depend on the selection goal. If the objective is to discard animals with clinical mastitis problems, CM could be a good definition. However, in the context of selection for overall profitability, because of the antagonistic relationship between milk production and mastitis resistance, considering the number of events (NCM) led to a greater ability to predict future costs associated with mastitis. Then NCM could be more useful for discarding animals with more than, for example, 2 episodes per lactation. Further research should consider the relative economic importance of NCM and production traits, and it is important to determine the extent to which increased milk revenue compensates for additional mastitis costs.

ACKNOWLEDGMENTS

The authors acknowledge support of the farmers involved in the Ministerio de Industria, Comercio y Turismo (MICT) program (PDT 2008-0278), as well as the Holstein Associations of Navarra (AFNA), Girona (AFRIGI), and Basque Autonomous Region (EFRIFE) for their cooperation. K. A. Weigel acknowledges support from the National Association of Animal Breeders (Columbia, MO).

REFERENCES

- Bar, D., Y. T. Gröhn, G. Bennett, R. N. Gonzalez, J. A. Hertl, H. F. Schulte, L. W. Tauer, F. L. Welcome, and Y. H. Schukken. 2007. Effect of repeated episodes of generic clinical mastitis on milk yield in dairy cows. *J. Dairy Sci.* 90:4643–4653.
- Bar, D., L. W. Tauer, G. Bennett, R. N. Gonzalez, J. A. Hertl, Y. H. Schukken, H. F. Schulte, F. L. Welcome, and Y. T. Gröhn. 2008. The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. *J. Dairy Sci.* 91:2205–2214.
- Bates, D., and D. Sarkar. 2005. *lme4*: Linear mixed-effects model using S4 class. <http://lme4.r-forge.r-project.org/> Accessed Jul. 8, 2008.
- Bates, D., and A. I. Vazquez. 2008. Pedigreemm: Pedigree-based mixed-effects models. <http://pedigreemm.r-forge.r-project.org/> Accessed Jul. 8, 2008.
- Carlén, E. 2008. Genetic evaluation of clinical mastitis in dairy cattle. PhD Diss. Dept. Anim. Breed. Genet., SLU (Swedish University of Agricultural Sciences). Acta Universitatis Agriculturae Sueciae vol. 2008:63. <http://diss-epsilon.slu.se/archive/00001836/> Accessed Nov. 18, 2008.
- Carlén, E., M. del P. Schneider, and E. Strandberg. 2005. Comparison between linear models and survival analysis for genetic evaluation of clinical mastitis in dairy cattle. *J. Dairy Sci.* 88:797–803.
- Efron, B., and R. J. Tibshirani. 1993. Cross-validation and other estimates of prediction error. Pages 237–255 in *An Introduction to the Bootstrap*. Chapman and Hall, New York, NY.
- Fitzmaurice, G. M., S. R. Lipsitz, and J. G. Ibrahim. 2007. A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics* 63:942–946.
- Foulley, J. L., D. Gianola, and S. Im. 1987. Genetic evaluation for traits distributed as Poisson binomial with reference to reproductive characters. *Theor. Appl. Genet.* 73:870–877.

- Guard, C. 2008. The costs of common diseases of dairy cattle. Proc. Central Vet. Conf. West, San Diego, CA. <http://veterinarycalendar.dvm360.com/avhc/Veterinary+Food+Animal/The-costs-of-common-diseases-of-dairy-cattle-Proce/ArticleStandard/Article/detail/586496> Accessed April 13, 2009.
- Hansen, M., M. S. Lund, M. K. Sorensen, and L. G. Christensen. 2002. Genetic parameters of dairy character, protein yield, clinical mastitis, and other diseases in the Danish Holstein cattle. *J. Dairy Sci.* 85:445–452.
- Heringstad, B., Y. M. Chang, D. Gianola, and G. Klemetsdal. 2004. Multivariate threshold model analysis of clinical mastitis in multiparous Norwegian dairy cattle. *J. Dairy Sci.* 87:3038–3046.
- Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal, and K. A. Weigel. 2003. Genetic change for clinical mastitis in Norwegian cattle: A threshold model analysis. *J. Dairy Sci.* 86:369–375.
- Hinrichs, D., E. Stamer, W. Junge, and E. Kalm. 2005. Genetic analysis of mastitis data using threshold models and genetic correlation with production traits. *J. Dairy Sci.* 88:2260–2268.
- Hortet, P., and H. Seegers. 1998. Loss in milk yield and related composition changes resulting from clinical mastitis in dairy cows. *Prev. Vet. Med.* 37:1–20.
- Interbull Center. 2008. Interbull routine genetic evaluation for udder health traits, August 2008. <http://www-interbull.slu.se/udder/framseida-udder.htm> Accessed Nov. 17, 2008.
- Ministerio de Medio Ambiente, Medio Rural y Marino, España. 2008. Informe quincenal situación del sector lácteo en España. Semanas 40 y 41 (del 29 de Sept. al 12 de Oct. de 2008). <http://www.mapa.es/app/vocwai/ListadoDocumentos.aspx?tg=informes&sec=let&lng=es> Accessed Oct. 30, 2008.
- Miglior, F., B. L. Muir, and B. J. Van Doormaal. 2005. Selection indices in Holstein cattle of various countries. *J. Dairy Sci.* 88:1255–1263.
- Negussie, E., I. Strandén, and E. A. Mäntysaari. 2008. Genetic analysis of liability to clinical mastitis, with somatic cell score and production traits using bivariate threshold-linear and linear-linear models. *Livest. Sci.* 117:52–59.
- Perez-Cabal, M. A., S. Yaici, and R. Alenda. 2008. Clinical mastitis in Spanish dairy cows: incidence and costs. *Span. J. Agric. Res.* 6:615–622.
- R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rauw, W. M., E. Kanis, E. N. Noordhuizen-Stassen, and F. J. Gromers. 1998. Undesirable side effects of selection for high production efficiency in farm animals: A review. *Livest. Prod. Sci.* 56:15–33.
- Rodrigues-Motta, M., D. Gianola, B. Heringstad, G. J. M. Rosa, and Y. M. Chang. 2007. A zero-inflated Poisson model for genetic analysis of the number of mastitis cases in Norwegian Red cows. *J. Dairy Sci.* 90:5306–5315.
- Shim, E. H., R. D. Shanks, and D. E. Morin. 2004. Milk loss and treatment costs associated with two treatment protocols for clinical mastitis in dairy cows. *J. Dairy Sci.* 87:2702–2708.
- Shook, G. E. 1989. Selection for disease resistance. *J. Dairy Sci.* 72:1349–1362.
- Shook, G. E. 2006. Major advances in determining appropriate selection goals. *J. Dairy Sci.* 89:1349–1361.
- Vazquez, A. I. 2007. Analysis of number of episodes of clinical mastitis in Norwegian red and Holstein cows with Poisson and categorical data mixed models. MS Thesis, Univ. Wisconsin, Madison.
- Vazquez, A. I., D. Gianola, D. Bates, K. A. Weigel, and B. Heringstad. 2009. Assessment of Poisson, logit and linear models for genetic analysis of clinical mastitis in Norwegian Red cows. *J. Dairy Sci.* 92:739–748.
- Wolfová, M., M. Štípková, and J. Wolf. 2006. Incidence and economics of clinical mastitis in five Holstein herds in the Czech Republic. *Prev. Vet. Med.* 77:48–64.
- Zwald, N. R., K. A. Weigel, Y. M. Chang, R. D. Welper, and J. S. Clay. 2006. Genetic analysis of clinical mastitis data from on-farm management software using threshold models. *J. Dairy Sci.* 89:330–336.

APPENDIX

Calculation of Mastitis Cost

Mastitis costs were obtained for each lactation with CM equal to 1 or NCM greater than or equal to 1. Mastitis cost for a healthy lactation was \$0. From information regarding the actual treatment applied to each case of clinical mastitis, cost on a lactation level was computed as:

$$mastc_{ijt} = \sum_{t=1}^5 (treatment_{ijt} + vet_{ijt} + labor_{ijt}) + red_production_{ijt},$$

where $mastc_{ijt}$ is the total mastitis cost in the j th lactation of the i th cow with t cases; $treatment_{ijt}$ is the cost of drugs applied to the t th case in the j th lactation of the i th cow; vet_{ijt} is the veterinary fee per case; $labor_{ijt}$ is the farmer's extra time dedicated to treating each case; and $red_production_{ijt}$ is the total cost attributable to lower production per lactation of a diseased cow with t cases relative to a healthy cow. Some other important costs were not considered in this study (such as an increased risk of other diseases) because, to date, little information is available regarding other diseases in this population. The exchange rate for costs attributable to treatment and the reduction of production was US\$1.51/€, whereas veterinary and labor costs were computed directly in US dollars following Guard (2008), as explained below.

Treatment Costs

Treatment costs were obtained from Perez-Cabal et al. (2008), including drugs and discarded milk costs. Drug costs per case were obtained as:

$$drugs = P_{unit} \cdot Units \cdot treatdays,$$

where P_{unit} is the cost per unit applied of each product, estimated from average prices for drugs; $Units$ is the daily dose; and $treatdays$ is the treatment period (in days) for each drug, assuming the standard protocols recommended by the veterinarians and the manufacturers.

Discarded milk costs per case were estimated as follows:

$$discMilk = milk \cdot P_{milk} \cdot (treatdays + withdrawal),$$

where $milk$ is daily milk production (in kg) obtained as explained below; P_{milk} is average milk price in 2008

(\$0.62/kg; Ministerio de Medio Ambiente, Medio Rural y Marino, España, 2008); *treatdays* is treatment period (in days) for each drug; *withdrawal* is the number of withdrawal days after the last dose, following the manufacturer's recommendations.

Daily milk production at the onset (diagnosis) of each mastitis event was estimated using monthly milk yield records as follows:

- If the diagnosis occurred between calving and the first monthly test, average daily production was estimated as production at the first test.
- If the diagnosis occurred between 2 moly tests, the period was split into 2 equal intervals, and daily production was estimated as milk yield at the midpoint of the corresponding interval.
- If the diagnosis occurred after the last monthly test, one of the following situations was considered:
- The dry day of the lactation was known. If the diagnosis occurred before the dry date, daily production was estimated as the milk production at the last test (zero otherwise).
- The dry day of the lactation was unknown. If the diagnosis occurred before 270 DIM, daily production was estimated as milk production at the last test (zero otherwise).

Veterinary and Labor Costs

Additional labor costs and veterinary fees were calculated based on the number of cases and costs per case reported in Guard (2008). Although those costs are from US milking cows, the farmers involved in this study considered them applicable to the Spanish scenario as well. Thus, labor costs were calculated based on an average rate of \$15 per hour, 10 min of additional labor per case, and the actual length of the treatment period (in days). Veterinary fees were considered as \$2 per case.

Milk Loss Costs

A linear mixed model was fitted to estimate the reduction in total milk yield per lactation, using the *lmer* procedure in R (Bates and Sarkar, 2005), depending on the number of cases of mastitis per lactation:

$$\text{milk}_{ijklm} = a_0 + a_1 \text{DIM}_{ijklm} + \text{NCM}_i + S_j + L_k + \text{hy}_l + \text{cow}_m + \varepsilon_{ijklm},$$

where milk_{ijklm} is the actual milk production in the k th lactation of the m th cow; a_0 is the average milk yield per lactation; DIM_{ijklm} is a covariate corresponding to total DIM of the k th lactation of the m th cow and a_1 is a regression coefficient; NCM_i is the fixed effect number of cases of mastitis (5 levels) in the k th lactation of the m th cow; S_j is the fixed effect of the j th season of calving (4 levels); L_k is the fixed effect of the k th lactation (7 levels); hy_l is the random effect of the l th herd-year of calving (92 levels); cow_m is the random effect of the m th cow (3,261 cows), and ε_{ijklm} is a random residual. The distribution of the random effects was as follows:

$$\begin{pmatrix} \mathbf{hy} \\ \mathbf{cow} \end{pmatrix} \sim N \left[\mathbf{0}, \begin{pmatrix} \mathbf{I}\sigma_{hy}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_{cow}^2 \end{pmatrix} \right],$$

where $\mathbf{hy}$ and $\mathbf{cow}$ were the vectors of herd-year of calving and cow effects, respectively, and σ_{hy}^2 and σ_{cow}^2 are the corresponding variances, estimated using REML; and $\mathbf{I}$ is an identity matrix of appropriate order.

The reduced milk yield per lactation was then considered as the estimated difference between a healthy lactation and NCM greater than zero. For example, with the reference of milk yield equal to zero for a healthy lactation ($\text{NCM} = 0$), if the estimation obtained for $\text{NCM} = 1$ is -100 kg, the average milk loss in a lactation with 1 episode of mastitis would be 100 kg.