Original article

An individual modelling tool for consecutive clinical mastitis during the same lactation in dairy cows: a method based on a survival model

Patrick GASQUI[a]*, Odile PONS[b], Jean-Baptiste COULON[c]

[a] Unité d'Epidémiologie Animale, INRA, 63122 Saint-Genès-Champanelle, France
[b] Unité de Biométrie, INRA, 78352 Jouy-en-Josas, France
[c] Unité de Recherche sur les Herbivores, INRA, 63122 Saint-Genès-Champanelle, France

Abstract – The high number of clinical mastitis recurring within the same lactation in dairy cows constitutes one of the factors of overdispersion in standard Poisson models. Our method, based on biological parameters, i.e., recurrence hazard in relation to udder exogenous infection (Rex) or recurrence hazard and rate in relation to endogenous infection (Ren), produced a model capable of integrating a possible change of state in the udder after clinical mastitis. This model was based on a study of the time intervals between successive clinical episodes, both types of risk being considered in the form of a distribution mixture in the survival model. The modelling tool allowed to determine the factors that specifically act on either one of the potential risks and estimated the distribution of the number of clinical mastitis per lactation, as well as the distribution of when mastitis occurs. Estimation results obtained by this method in an experimental herd were compared with those from more classical models with or without random individual effects. The distribution of the number of mastitis per lactation estimated by our method was well-fitted to the data and the method identified variation factors which were relatively standard in this type of study: lactation number, lactation stage and calving month. Prediction results obtained in another experimental herd with more recent data without parameter re-estimation demonstrated the adequacy of the model in fitting observed data. This modelling method based on biological parameters in a mixture of survival distributions was interesting to model clinical mastitis recurring within the same lactation. However in the future it will also be important to integrate the possible relationship between successive lactations and to apply this model to other types of farming systems.

clinical mastitis / individual model / dairy cow / survival model / recurrence

* Correspondence and reprints
Tel.: (33) 4 73 62 42 65; fax: (33) 4 73 62 45 48; e-mail: Patrick.Gasqui@clermont.inra.fr

Résumé – Un outil de modélisation individuelle des mammites cliniques successives au sein d'une lactation chez la vache laitière : une méthode basée sur un modèle de survie. Les nombres élevés de mammites cliniques successives au sein d'une lactation chez des vaches laitières constitue un des facteurs de sur-dispersion dans les modèles de Poisson usuels. Avec notre approche fondée sur des paramètres biologiques : risque de récidive lié à une infection exogène de la mamelle (Rex), risque et taux de récurrence liés à une infection endogène (Ren), on obtient un modèle apte à prendre en compte un changement d'état de la mamelle suite à une mammite clinique. Ce modèle est fondé sur l'étude des intervalles de temps séparant les événements cliniques successifs, les deux types de risque étant pris en compte sous la forme d'un mélange de distribution dans le cadre des modèles de survie. L'outil de modélisation permet la mise en évidence de facteurs agissant spécifiquement sur l'un ou l'autre des risques potentiels, et permet d'estimer la distribution du nombre de mammites cliniques par lactation ainsi que la distribution de leurs instants d'occurrence. Des résultats d'estimation obtenus à l'aide de cette approche dans un domaine expérimental sont comparés aux résultats issus de modèles plus classiques avec ou sans effet individu aléatoire. Notre approche montre une bonne adéquation de la distribution estimée du nombre de mammites par lactation à celle observée, et la mise en évidence de facteurs de variation relativement classiques dans ce type d'étude : rang de lactation, stade de lactation et mois de vêlage. Des résultats de prédiction obtenus dans un autre domaine expérimental avec des données plus récentes sans ré-estimations des paramètres montrent l'adéquation du modèle pour ajuster des données observées. Cette approche de modélisation fondée sur des paramètres biologiques dans le cadre d'un mélange de distributions de survie, semble intéressant pour modéliser les mammites cliniques au sein d'une même lactation. Cependant il serait important d'intégrer dans le futur une possible relation entre lactations successives, et d'appliquer le modèle à d'autres types de système d'élevage.

mammite clinique / modèle individuel / vache laitière / modèle de survie / récurrence

1. INTRODUCTION

Mastitis is still the most frequent health disorder in dairy cattle [7]. It is also the disease with the most deleterious consequences on milk quality, hence on economics [22]. The problem has been taken seriously enough for the dairy industry to envisage systematic recording of clinical mastitis cases on dairy cattle farms [24, 28]. Models for clinical mastitis occurrences are in this case valuable tools for herd management, in that they allow to study and rank the potential influencing factors [32]. Tools that simulate clinical mastitis occurrence are emerging, based on results from various earlier studies. They are intended to assist breeders in the decision making process while focussing on the economic consequences of the disease [3, 4, 23, 43]. The elaboration of models that (1) estimate the effects of various individual and environmental factors and (2) predict clinical mastitis occurrence while considering the effects of these factors, is confronted with a number of problems. These problems are now clearly identified and linked, in particular, to the relatedness of successive mastitis during lactation, to the relatedness of successive lactation during a cow's career, to the relatedness of cows in a herd and to the association of clinical mastitis with other diseases that may occur, in particular during the peri-calving period. To characterize the relatedness of successive mastitis at herd level, some authors have proposed to calculate a "reoccurrence rate" [38] (total number of clinical mastitis cases per number of lactations with at least one case), and a "lactational incidence risk" [8] (number of lactations with at least one occurrence of the disease of interest divided by the total number of lactations at risk).

With this prospect to estimate and predict, the classic approach generally consists in fitting to a Generalized Linear Model (GLM) [33]. However, adjustments using these models often elicit overdispersion, considered by many authors as a measure of deviation from baseline independence

hypotheses [11, 14–16, 34–36, 42]. In practice, various approaches have been proposed; the most widely used being through a negative binomial distribution instead of a Poisson distribution [42, 44]. Some authors have proposed using the Generalized Estimating Equations method (GEE) [13] to estimate model parameters with inter-response correlation within a group. But parameters from these models are difficult to interpret on biological terms and the models cannot relate successive mastitis occurrences during the same lactation. An alternative method classically consists in adding a random individual effect to the model, which takes into account all non-detectable variation factors that were omitted in the model, as well as a possible relationship between consecutive events [11]. This method presents the advantage of providing a model which ensures correct predictions but which on the other hand can hardly be generalised and used with other data and does not provide any practical explanation for the possible relationship between consecutive events. This drawback is also found in multivariate survival models, which generalise the Cox model (proportional risk model) [30] by including random effects [10]. A more recent approach based on multistate models takes the relationship between consecutive events into account by considering sub-clinical or clinical udder infection states as opposed to the healthy state of udders prone to becoming infected and by describing the probabilities of changing from one state to the other in a simulation framework [2, 5]; indeed, few data are available to describe the evolution of these various statuses with time in dairy cows. Another method consists in explaining the relationship between consecutive clinical mastitis using differentiated parameters that express recurrence in relation to exogenous infection (Rex) and recurrence in relation to endogenous infection (Ren) [17, 19]. This model also considers the relative effectiveness of the treatments applied after clinical mastitis in the course of lactation or at the time of drying

off, with such treatments not necessarily warranting bacteriological cure of the udder [9]. For Oltenacu et al. [40], "rates of cure for clinical infections at each period and stage are: (a) during the dry period: 72% with strep, 51% with staph, and 65% with other types of organisms, (b) during the lactation period: 65% with strep, 45% with staph, and 60% with other types of organisms". These various approaches correspond to the current statement of veterinary epidemiologists, who are convinced that better knowledge and assessment of clinical mastitis condition in herds requires explicit consideration of the dependence between consecutive cases in assessment and prediction models [2, 5, 23, 29, 40].

The aim of this study was to propose, within the scope of survival models [39], an individual modelling tool that allows to test the effect of various potential factors through tests on the intrinsic parameters of the model, the Rex hazard and the Ren hazard and rate, in the absence of bacteriological test results. This model is based on the analysis of time intervals between detectable successive events occurring in each cow and considers the relationship between consecutive clinical mastitis within one lactation [19, 20]. The method can later be generalised and extended to the potential relations between consecutive lactations or even between animals of the same herd. This modelling method is based on parameters which can be estimated and easily interpreted epidemiologically. The result is a model which predicts the distribution of the number of events per lactation as well as the distribution of the occurrence times of the events.

2. MATERIALS AND METHODS

2.1. Data

The data used to illustrate the interest of this method and to validate the model of the

relationship between consecutive events during the same lactation came from two experimental farms of the Institut National de la Recherche Agronomique (INRA), France: Orcival provided data from 1979 to 1989, which helped develop the model in terms of parameter estimation, and Marcenat provided data for the 1997-98 campaign, which helped validate the model in terms of prediction. The data from the former farm had already been used to elaborate the initial model for clinical mastitis occurrence proposed by Lescourret et al. [32]. Husbandry conditions on the Orcival farm were relatively homogeneous during the experimental period. A milking parlour was used, the animals spent 5 months in untethered stables (1st November to 31st March) and the rest of the time on a pasture. Three breeds were present on the farm: Holstein (HO), Holstein-French Friesian crossbreed (HO*FF) and the Montbéliarde-Pie Noire crossbreed (MO*PN). On the Marcenat farm, the husbandry conditions were relatively close to those of the former: a milking parlour was used, the animals spent 5 months in tethered stables (1st November to 31st March) and the rest of the time on the pasture. Three breeds were present on the farm: Holstein, Montbéliarde and Tarine, which led to the validation of the model using more recent data.

2.1.1. Selecting cows and lactations

The cows and lactations that allowed to study the relationship between consecutive mastitis during the same lactation were selected on both farms using the same criteria. Lactations with a productive time situated symmetrically between 210 and 390 days on either side of the mean recorded duration of 303 days, and cows for which all lactations were recorded were included in the study. Of these lactations, only those lactations rating from 1 to 3 were retained; they corresponded to calving periods between November and February, that is the majority on these farms. We then

excluded all lactations where a disease affecting the udder (other than mastitis) was present at the time of calving, and known to be possibly associated with the occurrence of clinical mastitis, i.e., udder oedema or vitular hypocalcaemia [12, 27]. Other peri-calving diseases known to be possibly associated with clinical mastitis [21] were infrequent or even nonexistent during the subject lactations. For each cow, the lactations from the first (parity 1) to the first one presenting one or several clinical mastitis were retained. Therefore, at the most there was one lactation with at least one clinical mastitis per cow, with all intermediate lactations with no mastitis or peri-calving diseases possibly associated to clinical mastitis, from parity 1 to parity 3 maximum. The hypothesis of the independence of consecutive lactations of the same cow was then verified by ruling out the possible influence of other peri-calving diseases on the occurrence of the recorded clinical mastitis.

Keeping only the first three lactations of each cow in a study that lasted 10 years in the first farm restricted the number of cows, hence the number of concurrent lactations, significantly reducing the potential influence of the degree of possible dependence between cows of the same herd. At the same time it allowed to compare the first three lactations, which is common practice on traditional cattle farms. This hypothesis of independence of cows of the same herd was all the more justified here since clinical mastitis did not exhibit any epidemic character and since no mastitis peak cluster was noted within a few days in the experimental herd.

In this study, all clinical mastitis recorded and diagnosed from clinical signs (milk aspect, conditions of quarters or udders) were documented. All clinical mastitis were retained regardless of the time interval between two consecutive occurrences because the model was aimed at estimating the possible relationship between consecutive mastitis during the same lactation.

Therefore, the model construction data included a sample of 315 lactations involving 194 cows that suffered from 110 clinical mastitis (47 lactations with 1 mastitis, 15 with 2, 5 with 3, 3 with 4 and 1 with 6). Data for validation included 39 lactations with only one lactation per cow and 20 clinical mastitis (5 lactations with 1 mastitis, 4 with 2, 1 with 3 and 1 with 4). In this case the exclusion criteria (productive duration, lactation number, calving period, absence of disease other than mastitis associated to calving, preceding lactation free of clinical mastitis) retained 37% of the herd lactation. These data were later compared with those estimated by the model.

2.1.2. Selecting factors

In order to construct the model based on the data from the first farm, a number of potential and essentially individual factors were introduced in the model: a breed factor with the three breeds, a calving month factor with the 4 selected months and a parity factor from 1 to 3. A production potential factor (estimated from the initial production, i.e., the mean production of the 4th, 5th and 6th day of lactation [32]) was not introduced in the model, because its estimation value potentially depends in particular, on anterior mastitis events and in general on anterior udder infection [22]. The possible effects of milk yield were confused with lactation stage, parity and breed effects. However an initial production factor with 4 levels (less than 16 kg, 16 to 20 kg, 20 to 24 kg and more than 24 kg) is simply studied with fitting results. Conversely, the lactation stage appeared to be worth considering. In order to investigate the risk of udder infection during lactation, some authors divided the lactation into consecutive periods. The first three days following calving carry a greater risk of udder infection, with risk factors more closely related to calving itself and to the preceding period [45]. The lactation curve of a dairy cow increases greatly during the 4th and 5th weeks and then decreases with time until the end of the productive period. A second period was therefore selected, from day 4 to day 30 of lactation, taking into account the sharp increase in milk yield during that period. At the same time, that period made it possible to take into account the apparently greater risk of clinical mastitis occurrence during the first 30 days than beyond [45]. Likewise, considering the period during which the cows were in stables on that farm, it was assured that during these first 30 days of lactation (the first two periods) all cows were in the stables. The overall milk production time of the retained lactations was no less than 210 days, with a mean duration of about 300 days; a period ranging from 210 to 300 days was therefore selected. During this period, the number of cows present varied in time according to effective drying off. This period was also characterised by the simultaneous presence of cows in the stables or on the pasture. To characterise the end of milk yield and in view of the date when the cows returned to the stables on the farm (1st November), from day 360 to the end of production (day 390), all cows not yet dried-out were sent back to the stables. The intermediate period from day 300 to day 360 was thus also characterised by a variable number of cows and the simultaneous presence of cows in the stables or on the pasture. In contrast, from day 150 to day 210 of lactation, all cows were on the pasture. The remaining period from day 30 to day 150 of lactation was divided into two sub-periods of equal duration, that is from day 30 to day 90 and from day 90 to day 150, during which all cows were present, in both the stables and grazing. The productive duration of a lactation was thus divided into 8 consecutive periods: calving to day 3, day 4 to day 30, day 31 to day 90, day 91 to day 150, day 151 to day 210, day 211 to day 300, day 301 to day 360 and beyond day 361 until the ultimate dry-out date, no later than day 390. The durations of these periods were 3, 27, 60, 60, 60, 90, 60 and 30 days, respectively, for an overall productive period of 390 days.

In the model proposed, the following potential factors were studied: lactation number (1, 2 or 3), HO breed, HO*FF breed, or MO*PN breed, calving month (November, December, January or February) and lactation stage, with the productive duration divided into 8 consecutive periods from calving onwards.

2.2. Modelling the relationship between consecutive mastitis during the same lactation

The Poisson distribution has for a long time been identified as the one closest to the number of mastitis during each lactation [24] or at herd level [37]. To estimate the parameters of a model based on this distribution, the authors generally choose GLM models. This approach takes the relationship between consecutive events into account, in the form of an additional parameter, i.e., the dispersion of a fixed effect GLM approach or the variance of an individual random effect in a mixed GLM. However, in both cases, the parameter is difficult to reuse for other data and in practice raises estimation problems [11, 33]. To explain a possible relationship between consecutive events, we studied the time interval between two consecutive mastitis in the framework of the survival models. A model was first introduced by using an exponential distribution for time intervals between the events. It was then extended to a distribution mixture which involved the parameters that characterise the relationship between consecutive mastitis during the same lactation.

2.2.1. The independent event model: Poisson's model

In dairy cows during production, clinical mastitis is diagnosed during milking. For each milking a random variable is observed, which can only take two modes: an "infected" or "uninfected" udder. A cow is subject to a large number of consecutive milkings during one lactation. Under the hypothesis that milkings are unrelated to one another, it is natural to consider that the random variable "number of milkings producing a diagnosis of clinical mastitis during a production interval" has a binomial distribution. Since the number of milkings is high and the probability of an event is low, this binomial distribution may be approximated by a Poisson distribution. We considered the occurrence times of clinical mastitis in the course of one lactation, in a given cow. If n mastitis were diagnosed at times t_1, t_2, …, t_n, after the origin time t_0 (calving date), we denote the time intervals between consecutive times by $d_1 = t_1 - t_0$, $d_i = t_i - t_{i-1}$ with $i = 2, …, n$ and d^* denotes the time interval between t_n and the dry-out time t^*. When the number of events occurring within nonoverlapping time intervals are independent and have a Poisson distribution, the random variable corresponding to the intervals between two consecutive events are independent and exponentially distributed [25, 26].

Assuming that the durations between mastitis are independent and have the same distribution, the likelihood can be written: $L = f(d_1) f(d_2) … f(d_n) S(d^*)$, where f is the density of the interval duration between mastitis, S is the associated survival function and t^* is considered as a right-censoring time. Thus the likelihood takes the whole productive duration into account, whether clinical mastitis was observed or not during that productive period. The probability that a cow suffers mastitis at moment d, with the condition that it had not suffered any until then, is the hazard function $h(d) = f(d)/S(d)$ for any $d \geq 0$. For an exponential distribution with parameter λ, the hazard function is constant and $h(d) = \lambda$ for any $d \geq 0$, $f_\lambda(d) = \lambda.\exp(-\lambda.d)$ and $S_\lambda(d) = \exp(-\lambda.d)$; the parameter λ is interpreted as the average number of events per day if the intervals are expressed as days; $1/\lambda$ is then the average time interval (in days) between two consecutive events. For one lactation in a cow, let $N(t_0, t)$ be the number of mastitis occurring from t_0 up to a time t preceding

t^*. We first assume that N is a homogeneous Poisson process observed in the interval $[t_0, t^*]$, with a constant hazard λ. Then the random variable W counting the clinical mastitis until t^* has a Poisson distribution with parameter $(\lambda.(t^* - t_0))$, i.e. for every integer $w \geq 0$, $P(W = w) = \lfloor(\lambda.(t^* - t_0))^w/w!\rfloor \exp\{-\lambda.(t^* - t_0)\}$. This is the Poisson distribution habitually used for the number of events occurring during a lactation in the GLM models. With an exponential distribution and a series of n^* unrelated lactations with different productive durations denoted $(t^*_j - t_{0j})$ for lactation j, the number of events W_j in lactation j has a Poisson distribution whose parameter is $(\lambda.(t^*_j - t_{0j}))$. If n^* independent lactations are considered globally, including n mastitis, the likelihood may be written:

$$L(\lambda) = \prod_{i=1}^{n} f_\lambda(d_i) \cdot \prod_{j=1}^{n^*} S_\lambda(d^*_j) = \lambda^n \cdot \exp\left\{-\lambda \sum_{j=1}^{n^*}(t^*_j - t_{0j})\right\}. \quad (1)$$

This expression leads to an estimator $\hat{\lambda} = n / \sum_{j=1}^{n^*}(t^*_j - t_{0j})$ of the parameter, obtained by maximising the likelihood with respect to λ; this estimator depends on the productive durations $(t^*_j - t_{0j})$ actually observed for the n^* lactations. So the number of events $W = \sum_{j=1}^{n^*} W_j$ for all the variable-duration lactations does not have a Poisson distribution and a dispersion appears as compared to a classical Poisson variable.

Let us consider the case where the hazard is constant throughout each lactation but different from one lactation to the other. This is the case when the individual characteristics of cows, for example parity, alter the risk of clinical mastitis. For lactation j, the mastitis variable count W_j has a Poisson distribution with parameter $(\lambda_j \cdot (t^*_j - t_{0j}))$. There again the number of events W for all lactations with variable durations and hazards is different from a Poisson variable, with an overdispersion with respect to the standard Poisson variable.

With the GLMs, such models can take into account individual or environmental fixed effect factors which characterise the lactations, and include the observed lactation duration in the form of a covariate (offset) as well as a random-effect cow factor. This type of model was used to compare the results obtained with this GLM method (mixed model) and those defined hereafter with a survival method based on a likelihood that generalised $L(\lambda)$ (Eq. (1)) and also integrated a lactation stage factor and a relation between consecutive events. The mixed model used in the GLM approach (MM model) was written:

$$g(E(W)) = X \cdot \theta + \text{offset}(\log(t^* - t_0)) + a,$$

where $E(W)$ is the mean number of mastitis in a lactation according to a Poisson distribution, g is the associated canonical link function (log function), X is the incidence matrix, θ is the vector of the parameters to be estimated, which contains the fixed effects of the factors included in the model (3-mode parity, 3-mode breed and 4-mode calving month), the offset covariate offset($\log(t^* - t_0)$) expressing the lactation time logarithm and a random variable "a" with a Gaussian distribution of mean zero and variance σ_a^2 characterising the random cow factor [41].

During one lactation, the hazard parameter can also be considered as non constant because of physiological factors (higher risk at the very beginning of lactation and variable henceforth according to the distance from the calving date) or environmental factors (higher risk in stables than on the pasture). Given that the risk is simply constant throughout the consecutive periods within lactation, a Poisson process homogeneous for fixed intervals is stipulated. These intervals may be constant from one lactation to the other (post-calving periods and production) or variable (e.g., according to the duration of stables or grazing periods). Consider a productive duration divided in K consecutive periods $\lfloor s^*_{k-1}, s^*_k \rfloor$, for each

k period varying from 1 to K where the period ending time is s^*_k, the period length is $u_k = s^*_k - s^*_{k-1}$, the constant hazard associated to that period k is λ_k, and the number of events occurring during that period is W_k. Each W_k variable has a Poisson distribution with parameter $(\lambda_k \cdot u_k)$. If the variables W_k are independent and if $W = \Sigma_k W_k$ is the number of events occurring throughout the production time $\sum_{k=1}^{K} u_k$ of a lactation, the variable W does not have a Poisson distribution; therefore an overdispersion is also present. With GLM and assuming that consecutive mastitis within a period or with other consecutive periods of the same lactation are unrelated, a model can be defined which takes a multi-period lactation stage, individual and environmental factors as well as the actual lactation duration in the form of a covariate (an offset) into account. This model was used to compare a GLM approach and a survival approach integrating an additional relationship between consecutive events. The GLM model used (MP model) is therefore:

$$g(E(W_k)) = X \cdot \theta + \text{offset}(\log(u_k)),$$

where "W_k" is the number of mastitis during period k, a number that follows a Poisson distribution, g is the associated canonical link function (the log function), X is the incidence matrix, θ the vector of the parameters to be estimated, which includes the fixed effects of the factors introduced in the model (8-mode lactation stage, 3-mode parity, 3-mode breed and 4-mode calving month), and a covariable offset($\log(u_k)$) expressing the logarithm of the period duration. From estimators of the parameters of this model, the probability of recording w mastitis over the lactation can be expressed according to the probabilities of having recorded w_k mastitis during the period k. Thus the distribution of the number of mastitis cases per lactation can be estimated.

These three phenomena, i.e., productive duration, variable hazard from one lactation to the other, and variable hazard from a period to another within the same lactation, accounted for some of the overdispersion found when adjusting the observed number of mastitis with a Poisson distribution. The survival model took these three aspects into account: the observed duration and the individual and husbandry characteristics responsible for the differences in risk levels. It also integrated other biological elements whereby at the herd level a basic Poisson distribution was more remote and overdispersion appeared. This was due to considering hazards that vary with time (IN or OUT of the stables at different times in relation to calving date, according to cows, for example) and above all considering links between consecutive mastitis, a phenomenon that is the subject of the next section.

2.2.2. The model for the relationship between consecutive mastitis during the same lactation

The relationship between consecutive mastitis during the same lactation appears to be the key factor in terms of overdispersion. This relationship cannot be considered by modelling mastitis count during one lactation, which would correspond to the GLM method. This GLM postulates that consecutive mastitis and lactations are independent. The GEE method [13] still postulates that consecutive mastitis during the same lactation are unrelated but it allows to introduce a dependence between consecutive lactations. The relationship that exists between consecutive mastitis during the same lactation may, in contrast, be considered explicitly within the framework of survival models.

We assumed that the relationship between consecutive clinical mastitis was established as follows: at the outset of clinical mastitis, the cow was exposed to the risk of having mastitis again with a hazard λ in a proportion $(1 - p)$ of cases, at a level identical to that to which it was submitted before mastitis, whereas in a proportion p of cases the cow would be subjected to a high level of risk with a hazard λ_r, higher

than before. The hazard λ expresses a recurrent risk linked to exogenous infection (Rex) by new germs newly entering the udder, exclusive of the germs responsible for the preceding mastitis. The hazard λ_r expresses a recurrent risk linked to endogenous infection (Ren) induced for example by the persistence in the udder of the germs from the preceding mastitis. This model describes, for example, a situation where the treatment carried out during acute clinical mastitis obliterates mastitis symptoms without resulting in total eradication of the germs from the udder [46], in a proportion p of cases. It was also adapted to the case of clinical mastitis affecting a quarter; other quarters then run a much greater risk of suffering mastitis later than they would have, had no quarter of the udder ever been infected [1]. It was adapted in fact in all cases where there was a possible state change in the udder after clinical mastitis. In that case, the risk for a cow to develop mastitis within a relatively short time was higher than if the mastitis had been fully cured and the germs totally eradicated from the udder. In this case the hazard λ_r can also be considered in a first approach as constant throughout all consecutive periods of a lactation as well as during all lactations, because the risk was due to a new persistent state of the udder. Proportion p is a Ren index. So for each clinical mastitis, further risk was reassessed by studying the time intervals between mastitis in a mixture distribution model. Thus a model was defined which explicitly considers the possible link between two consecutive events (MI model) within the survival model framework, with non-random effects. For this model, the likelihood formula is given in Appendix I.

For a series of n^* independent lactations, the overall likelihood is simply expressed as the product of the likelihood of each lactation, or: $L(\theta, \alpha, \beta) = \prod_{i=1}^{n^*} L_i$. The parameter θ is the vector of the model coefficients that describes the effects of the various factors, including the period factor, which influences the Rex hazard through a relation of the form $\lambda = \exp(X_1 \cdot \theta)$, where X_1 is the associated incidence vector containing, in particular, the period factor. The exponential relation warrants that the hazard calculated this way be positive. The parameter β is the vector of the model coefficients related to the various factors that weight on the Ren hazard through the similar relation $\lambda_r = \exp(X_2 \cdot \beta)$, where X_2 is the associated incidence vector. The parameter α is the vector of the model coefficients that reflect the effects of the various factors weighting on the Ren rate through the relation: $p = \exp(X_3 \cdot \alpha)/[1 + \exp(X_3 \cdot \alpha)]$, where X_3 is the associated incidence vector. This latter relation warrants a p value between 0 and 1, which clearly fits the variation range of probability. This total likelihood therefore solely depends on unknown coefficients θ, β and α so defined. Maximising the full likelihood according to the unknown coefficients of the model will provide estimates of these coefficients.

Given the parameters of the model, Rex hazard and Ren hazard and rate, the probability law of the random number of mastitis over any lactation duration from calving can be expressed; the probability formulas are given in Appendix II. Again in this model, the clinical mastitis count over the whole lactation does not have a Poisson distribution.

The interest of the proposed approach was that it also provides an estimator of the distribution function for occurrence time of the wth mastitis from the calving date, $P(T_w \leq t)$ for any occurrence time $t > t_0$, based on parameter estimators. Indeed, $P(T_w \leq t) = P(N(t_0,t) \geq w) = 1 - P(N(t_0,t) < w)$ can be deducted from the recurrence formulas of Appendix II.

2.3. Estimation of the model parameters including the relationship between consecutive mastitis

In the MI model, adjustments were performed using a function of the Fortran

mathematical library "Numerical Algorithms Group" (NAG Ltd, Oxford, UK, Mark 16 under Unix) which uses an explicit likelihood function and its first two derivatives with respect to the paramaters obtained by finite differences. The variance-covariance matrix estimate was obtained from the estimate of the inverse Fisher information matrix, the Hessian matrix being estimated by a NAG function by finite differences. The need to obtain a Ren hazard λ_r rigorously greater than the Rex hazard λ_k for any lactation period k ($k = 1, ..., K$) can sometimes require the use of a maximising function under parameter inequality constraints.

The distribution of coefficient estimators and those of the linear combinations of the coefficients constructed by maximum likelihood in a regular model, converge toward Gaussian distributions under the hypothesis that the model can be identified [31]. These limits provide confidence intervals associated with coefficient estimates or their linear combinations. Confidence intervals for the model parameters (Rex hazard and Ren hazard and rate) can be deduced by transformation from estimates of the associated coefficients; these confidence intervals are classically asymmetrical.

Sub-model tests can be performed by using the likelihood ratio statistics

$$\Lambda = -2 \cdot \log(L_0 / L_1)$$

which follows a χ^2 with $(q_0 - q_1)$ degrees of freedom if L_0 is the likelihood of the general model with q_0 coefficients and if L_1 is the likelihood of the sub-model with q_1 coefficients. These statistics allow to test the effects of the various individual factors while maintaining coherence with the estimator of the maximum likelihood used to estimate coefficients and hence the parameters of each model.

The MP model parameters were estimated with the "glm()" function of the "Splus" software (Statistical Science, Inc., Seattle, USA, Version 3.4 under Unix). With the MM model, the "Bayesian inference Using Gibbs Sampling" software was used (BUGS, MRC Biostatistics Unit, Cambridge, UK, Version 0.5 under Unix).

2.4. Validation of a model including the relationship between consecutive mastitis

Full model validation is achieved by using martingale residuals classically used in survival models [6, 18]. In that approach, the observation is a duration which ends either in mastitis or at the mastitis-free end of lactation (right-censorship) for each lactation recorded. The martingale residuals are defined in Appendix III. When the observation is a censorship, the residual is strictly negative. Examining martingale residuals according to the various individual factor values or the rank of the observations in relation with time provides a validation of the model if the mean of the residuals is zero (no systematic bias) and if there are no isolated points.

For a lactation j such as $j = 1, ..., n^*$, with an actual productive duration ($t^*_j - t_{0j}$), if W_j counts the clinical mastitis of this jth lactation, the result is:

$$P(W_j = w) = P(N(t_{0j}, t^*_j) = w)$$

for any $w \geq 0$. Hypothesising that cows and consecutive lactations in the same cow are unrelated, the distribution estimate for the number of mastitis per lactation is expressed as:

$$P(W = w) = \frac{1}{n^*} \cdot \sum_{j=1}^{n^*} P\left(W_j = w\right)$$

for any $w \geq 0$, and should be compared with the distribution actually observed.

2.5. Comparison between the three models MI, MP and MM

The MP model is a GLM model and is fit using iterative reweighted least squares (IRLS). This model is defined with 15 parameters. The MI model is a mixture survival model and is fit using maximum likelihood. This model is defined with

7 parameters. The difference between the MI and MP model parameters is the use of Ren parameters (hazard and rate) in the MI model. The MM model is a mixed GLM model and is fit by bayesian inference using Gibbs sampling. This model is defined by 201 parameters. The difference between MI and MM model parameters is the use of Ren and lactation stage factor parameters in the MI model, and the use of random cow factor parameters in the MM model.

The three models are not nested and fit using different methods; according to our knowledge, there is no statistical test that compares them. Likewise with a χ^2 test, it is not possible to compare the distribution of the number of clinical mastitis estimated by lactation for each model with the distribution observed, because the estimated parameter number is too big and a maximum likelihood is not always used. On the contrary, it is always possible to calculate a χ^2 value for each model. Also it is possible to break down this total χ^2 value into χ^2 values by mastitis number per lactation. These values permit to quantify the goodness of fit of a model, whatever its nature.

3. RESULTS

3.1. Estimating model parameters with the Orcival farm cows

In the GLM based on a homogeneous Poisson distribution, a 1.15 overdispersion was obtained. The "reoccurrence rate" was 1.55, and the "lactational incidence risk" was 22.5%. The survival model previously introduced was also used, taking into account the relations between consecutive clinical mastitis within the same lactation and the division of lactations into 8 successive episodes from calving, with more classical factors (parity, calving month and breed). In this full model, the martingale residuals were null on average, regardless of the modalities of potential factors, and

they varied between –1.0 and +1.0, thus validating the model.

From the full model, successive submodel tests (forward selection) at a confidence level of 95% led to considering the breed factor as non significant (0.173 P-value), and to summing up parity 2 and 3 and retaining only 3 consecutive periods "0 to 3 days", "4–150 days" and "151 days to end of lactation" (0.732 P-value). The modalities of the calving month factor were also gathered into two new modalities: "November or December" and "January or February" (0.620 P-value). In this model, the Rex hazard only depended on 3 consecutive periods during lactation, one 2-modality lactation stage factor (primiparous vs. parity 2 or 3 multiparous) and a 2-modality calving month factor (November or December vs. January or February). Ren hazard and rate were always considered as constant throughout the duration of a lactation for all selected cows, because there were not enough recurrent cases in the data used.

The estimates of the Rex hazard parameters obtained in the final model, as well as those of their standard deviations and 95% confidence interval are shown in Table I for calving in November or December and in Table II for calving in January or February. It was noted that the Rex hazard was 2.6 times lower for January or February calvings than for November or December

Table I. The estimates of the Rex parameters ($\times 10^{-5}$) obtained with the final model, their standard deviations in parentheses, and their 95% confidence interval, for calvings in November or December.

Parity	Period day 0 to 3	Period day 4 to 150	Period ≥ day 151
1	425 (196) (172; 1051)	120 (21) (85; 171)	21 (7) (11; 40)
2 or 3	908 (418) (368; 2236)	257 (45) (182; 363)	45 (14) (24; 83)

Parity	Period day 0 to 3	Period day 4 to 150	Period $\geq$ day 151
1	166 (96) (53; 515)	47 (18) (22; 101)	8 (4) (3; 21)
2 or 3	354 (197) (119; 1053)	100 (36) (49; 203)	18 (8) (7; 42)

calvings. This hazard was 2.1 times higher in multiparous than in primiparous cows. Lastly, it was 3.5 times higher during the first 3 days of lactation than during the following 147 days, and 20.2 times higher than at the end of lactation.

The Ren hazard was estimated as 0.02853 (0.00984), with a 95% confidence interval of (0.01451; 0.05608). The mean number of days between two mastitis, the second of which was recurrent of the first one, was thus estimated as 1/0.02853, that is 35 days or 5 weeks, with a 95% confidence interval of (18; 69) days or (2; 10) weeks. The Ren rate was estimated as 0.23 (0.06), with a 95% confidence interval of (0.14; 0.35). The distribution function estimates $P(T_1 \leq t)$ and $P(T_2 \leq t)$ obtained with the final model are displayed in Figure 1. These distributions permitted to calculate the distribution of the number of clinical mastitis estimated by the MI model (0, 1 and 2 and more) at any day after calving time. For example, it was possible to verify the values in Table III for the MI model and the 0, 1 and 2 and more cases, with these distributions at 390 days.

Distribution estimates of the number of mastitis for each lactation were also obtained by explicitly considering the relationship between consecutive events during the same lactation (MI model). The distribution of the observed number of mastitis was compared to the distribution of the number of
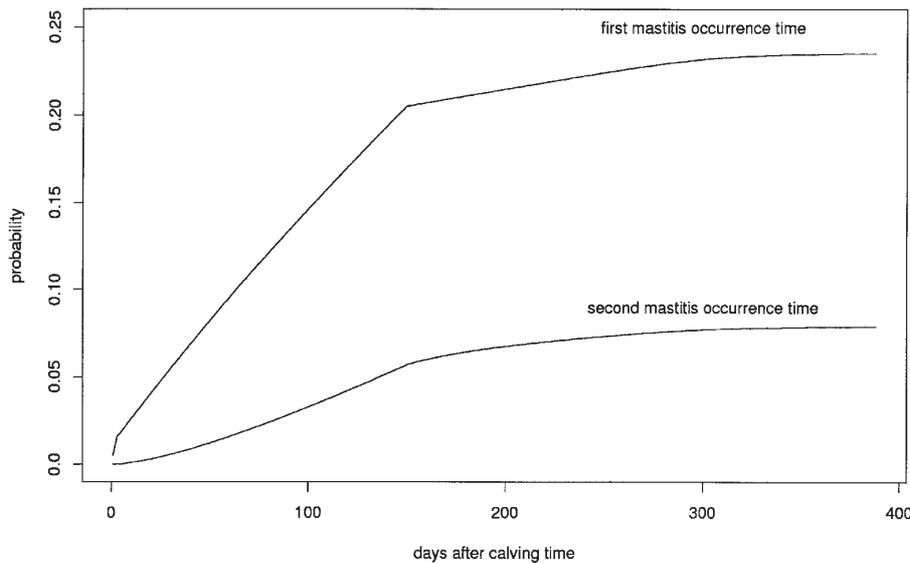
Figure 1. The distribution function estimates $P(T_1 \leq t)$ and $P(T_2 \leq t)$ obtained with the final model for the two first clinical mastitis.

Table III. Distribution of the number of clinical mastitis observed and estimated with three models, for 315 lactations.

Mastitis number per lactation	0	1	2	3 and more
Observed number	244	47	15	9
Estimated number with MI model	240	50	17	8
Estimated number with MP model	227	70	15	3
Estimated number with MM model	236	59	14	6

Table IV. For each model, χ^2 values calculated for each number of clinical mastitis per lactation, total χ^2 values and estimated parameter number.

	χ^2 for 0 mastitis	χ^2 for 1 mastitis	χ^2 for 2 mastitis	χ^2 for 3 and more mastitis	Total χ^2	Estimated parameter number
MI model	0.067	0.180	0.235	0.111	0.593	7
MP model	1.273	7.557	0.000	12.000	20.830	15
MM model	0.271	2.441	0.071	1.500	4.283	201

mastitis estimated by the fixed-effect GLM (MP model) and also to the distribution of the number of mastitis estimated by the mixed GLM, including a random individual effect in addition to the fixed effects (MM model). All these distributions were shown in Table III. The χ^2 values (by mastitis number and total) and estimated parameter number were shown in Table IV. It was thus noted that the approach chosen (MI model) provided a small total χ^2 value (0.593) in comparison with the MM model value (4.283) and the MP model value (20.83). In comparison with the MP model, the Ren parameters (hazard and rate) defined in the MI model seemed to improve the goodness of fit. This goodness of fit was observed for all classes for the MI model. In comparison with the MM model, the Ren parameters (hazard and rate) defined in the MI model seemed to improve goodness of fit less, but permitted to reduce the estimated parameter number (201 to 7). In terms of goodness of fit and estimated parameter number, the MI model seemed to be a good model.

Distribution estimates of the number of mastitis for each lactation and for initial production level were obtained by the MI model, and were compared to the distribution of the observed number in Table V. The χ^2 values (by mastitis number and total) were shown in Table VI. Goodness of fit was observed for more than 2 and more mastitis per lactation for all initial production levels. A little difference was observed between estimated and observed number for one mastitis per lactation, for less than 16 kg and more than 24 kg initial production levels (1.333 and 1.455 χ^2 values). But a χ^2

Table V. Distribution of the number of clinical mastitis estimated (observed) with the MI model by initial production level.

Initial production level	0	1	2	3 and more
Less than 16 kg	72 (75)	12 (8)	4 (4)	2 (2)
16 to 20 kg	69 (70)	13 (13)	4 (4)	2 (2)
20 to 24 kg	57 (61)	13 (11)	5 (4)	2 (2)
More than 24 kg	42 (38)	11 (15)	4 (3)	2 (3)

Table VI. χ^2 values calculated for each number of clinical mastitis per lactation and total χ^2 values by initial production level, with the MI model.

Initial production level	χ^2 for 0 mastitis	χ^2 for 1 mastitis	χ^2 for 2 mastitis	χ^2 for 3 and more mastitis	Total χ^2
Less than 16 kg	0.125	1.333	0.000	0.000	1.458
16 to 20 kg	0.014	0.000	0.000	0.000	0.014
20 to 24 kg	0.281	0.308	0.200	0.000	0.789
More than 24 kg	0.381	1.455	0.250	0.500	2.586

Table VII. Distribution of the number of clinical mastitis observed and estimated with the final model without Ren parameters, for 315 lactations.

Mastitis number per lactation	0	1	2	3 and more
Observed number	244	47	15	9
Estimated number with the MI model without Ren parameters	240	63	10	2

test realised with 0, 1 and 2 and more levels to have size of classes superior to 5, provided a 0.85 P-value with 8 degrees of freedom and 4.04 χ^2 value. The initial production factor was not statistically significant.

3.2. Predicting the distribution of the number of mastitis per lactation if there was no Ren, using the Orcival farm data

From the final model estimates, it was possible to predict the distribution of the number of mastitis that would have been observed if there had been no Ren, that is, if the treatments applied for clinical mastitis had been fully efficient against the germs present. This distribution was obtained by using the Rex hazard estimates previously obtained and setting the Ren hazard at a null value. This distribution, presented in Table VII, was compared to the observed distribution. It was then possible to note, for that farm, that there was no lactation with more than 2 clinical mastitis over a calving season in a 100-head population. The probability of lactations with 3 or more clinical mastitis was estimated as below 1% (0.63%).

3.3. Prediction involving the Marcenat farm cows, based on the preceding estimates of model parameters

From the hazard estimates based on the Orcival farm data and by using the Marcenat farm data factor modalities, the distribution of the number of mastitis per lactation was predicted with the MI model. This distribution was compared to the observed distribution shown in Table VIII. A χ^2 test

Table VIII. Distribution of the number of clinical mastitis observed and predicted on the Marcenat farm data (39 lactations) using parameter estimate values obtained with the final MI model and Orcival farm data.

Mastitis number per lactation	0	1	2	3 and more
Observed number	28	5	4	2
Predicted number	28	7	3	1

(1.904 χ^2 value and 0.59 P-value for 3 degrees of freedom) revealed the good prediction achieved for this second data set without having to re-estimate the model parameters.

4. DISCUSSION

Modelling of mastitis occurrence within a lactation for the number of cases and times of clinical mastitis is difficult because of the constraints of observational or experimental studies. As pointed out by Morse et al. [38], numerous data are required for studying uncommon events. The events should also be recorded similarly. Definitions also vary among authors. In our study, for each mastitic cow, all cases were considered, which led to an estimate of 1.55 for the "reoccurrence rate" of mastitis, and to an estimate of 22.5% for "lactational incidence risk". McMillan et al. [37] who also considered all cases, found 1.50, and Bigras-Poulin et al. [8] who considered cases separated by at least 10 days, found 1.47. The "lactational incidence risk" were respectively 12.3% and 25.0%. In this sense, the overdispersion we obtained was consistent with literature reports on clinical mastitis, even though this study was restricted to a single herd and the sampling procedure applied practically eliminated all links between consecutive lactations and those between cows of the same herd.

The MI model, by integrating the possible relationship between consecutive events, yielded the best fitting of the observed number of mastitis per lactation (total and individual χ^2 values) with a minimum estimated parameter number, among the proposed models. The consideration of this potential relationship appeared to be determinant with regard to the MP model, which only considered various periods within the same lactation. Considering both this dependence and the period factor clearly appeared to explain what was empirically expressed through the random individual factor in the mixed MM model. The advantage of the approach that involves a mixture distribution for modelling the relationship between consecutive events and excludes any random individual factor, is that it provides not only a prediction model that fits reality but also provides an explanatory model for successive occurrences of clinical mastitis within the same lactation in the same cow, including in particular the consideration of a possible change of state in the udder after clinical mastitis. This method therefore presents the advantage of considering a result already known, i.e., that only 50 to 80% of clinical mastitis are bacteriologically cured during lactation [46]. Of course, this percentage varies according to the type of germ involved, data typically not available on farms, but also according to the treatments used. For example, the relative effectiveness of the antibiotic treatments used during the productive period is a possible change of state in the udder after clinical mastitis.

Since this model is explanatory, it provides, at the same time, a model for the distribution of the occurrence times of the events according to their rank, ($P(T_1 \leq t)$, $P(T_2 \leq t)$, etc. With these distributions it is possible to determine a percentage of lactation with 0, 1 or 2 and more clinical mastitis at any day after calving.

As in all studies conducted so far on clinical mastitis, the confusion effect can be relatively significant between potentially interesting factors (in particular parity, milk yield potential and breed). Only a parity restricted to the primiparous vs. parity 2 or 3 multiparous alternative and a calving month effect restricted to the November or December vs. the January or February alternative could be clearly identified with the Rex hazard. These effects were revealed by considering the possible recurrences of clinical mastitis during the same lactation in the same animal. The advantage of such an approach was that it explored the potential and specific effect of a factor on either one of the model parameters: Rex hazard and Ren hazard and rate. The estimates obtained for these factors were

all the less biased and the values obtained can therefore be used more legitimately for other farms. Finally, the relatively good prediction values obtained by production potential level, factor not present in the model, appear to point to a negligible effect of milk yield on clinical mastitis with the data used.

The notion of Ren considered here at the udder level consisted in envisaging that a quarter affected by clinical mastitis increases the risk that other quarters would be later affected or that it would itself suffer another attack resulting from germ persistence within the quarter in question (through a lack of effectiveness of antibiotic treatment in eradicating all the infectious germs from the udder [46], hence through the bacteriological non-cure of the previously treated quarter, even if clinical symptoms were noted as disappearing). In the first case, it amounts to considering that the infected quarter more or less transmitted the infection to one or several other quarters, via the milking machine, for instance. In the second case, this amounts to considering that the antibiotic treatment did not totally eradicate the germs from the infected quarter. In the absence of the datum "quarter infected during a detected clinical mastitis", the two cases could not be separated. It is clear that Ren parameters correspond to a general state change of the udder after a clinical mastitis. Other reasons possibly lead to the modification of udder state: a deterioration of the udder by germs or a defective immune system at udder level, for example. This of course does not question the physiological independence of each udder quarter. It was noted that the Ren rate obtained with the Orcival farm data (0.23 with a 95% confidence interval of (0.14; 0.35)) appeared consistent with the few data found in the literature (average rate in the United Kingdom: 18.3% for example [29]), considering that in practice, the type of germ incriminated was usually not documented for clinical mastitis.

The Ren hazard of clinical mastitis was higher in early lactation and more so at the time of calving, and the hazard decreased during the productive period. This hazard decrease appeared not to be affected much by the marked increase in production recorded until the 5th or 6th production week. Conversely, the 150th day, when it was ascertained that all cows were in the pasture, appeared non-suppressible and clearly corresponded to a reduction of the clinical mastitis hazard. In contrast, because all calvings took place in the stables, the "stable/pasture" factor could not be investigated, that is during the first 30 days of lactation when calvings took place on the pasture.

The hazard of the first three days appeared relatively high, in relation to the subsequent hazards. This might mean that it took all that happened before calving into account, in particular what happened during the preceding dry period or even the one before for multiparous cows, or what happened before they calved for the first time in primiparous cows. These results were consistent with those obtained by Todhunter et al. [45], i.e., the hazard of udder infection during the first three days following calving depends on events occurring before calving, with the treatment applied at drying off after the previous lactation. This Rex hazard parameter also took some part of a possible relation between consecutive lactations into account. The implications of this would be that higher hazards in early lactation would be related to physiologic factors in reaction to mechanical milking or to calving, for instance, or even to infections caught at the end of the dry period or remnants of infection from the preceding lactation, rather than environmental factors in the wider sense of the term, such as husbandry practices in particular.

An interesting conclusion of this study is the fact that the levels of hazard obtained with data collected between 1979 and 1989 on one experimental farm were close to the hazard levels that could be measured with more recent data from another experimental farm of the same Auvergne region, collected during the 1997–1998 calving season.

With the model proposed, simultaneous estimation of Ren hazard and rate could prove difficult in practice, especially since populations are restricted in terms of lactation and clinical mastitis. This is a common problem with mixture distribution models when the number of potential factors is high. This is the most delicate point of the method. To correctly estimate Ren hazard and rate parameters, it is imperative to have sufficient data, i.e., a sufficient number of Ren mastitis. Despite this difficulty, the model appears relatively capable of adjusting the distribution of the number of mastitis in each lactation at the herd level by simply applying the lactation rank factor (limited to 1 vs. 2 or 3), to the consecutive periods factor (with only 3 periods) and above all to the consideration of a relation between consecutive clinical mastitis. Thus a relatively simple model is obtained where overdispersion no longer appears, neither on the construction nor on validation data. This was made possible through the survival model approach, which took the relevant biological hypotheses into account directly in the model. The fact of achieving relatively good adjustments with the hypotheses cited above appears consistent with a rather good independence of consecutive lactations in the same cow, with the sampling procedure chosen, and could justify the interest of this type of approach, whereas with more classical GLM or even GEE approaches, only a possible relation between lactations could have been considered. Conversely, the more common procedure is to introduce a random individual effect in a model to consider a possible dependence of events and factors that could not be a priori identified [11]. The method that is the more specific to the area studied is the modelling method proposed that considers the relationship between consecutive clinical mastitis during one lactation and does not have the same general aspect of the previous mixed models. In contrast, the model which includes this relationship will take the biological specificities of the farm into better consideration, and will better define easily interpretable parameters and provide good predictive adjustments.

The following determinant step will therefore be the integration of a possible relation between consecutive lactations and the potential influence of other pathologies during the pre-calving period, susceptible of being associated to clinical mastitis during lactation in this model and hence in the modelling tool. It will also be necessary to apply and to validate this modelling method to other types of farming systems with different types of mastitis problems and different factors.

ACKNOWLEDGEMENTS

APPENDIX I

Likelihood of mixture survival model with Ren parameters

Here we use the notations of Sections 2.2.1 and 2.2.2 and we assume that the Rex hazard is constant for the intervals $\rfloor s^*_{k-1}, s^*_k \lfloor$. For a given lactation, if $n > 0$ clinical mastitis are observed at times $t_1 < t_2 < \ldots < t_{n-1} < t_n$, so that $t_0 < t_1$ and $t_n < t^*$, the likelihood can be written:

$$L\left(\lambda_1, \ldots, \lambda_k, p, \lambda_r\right) = f\left(t_1 - t_0\right) \cdot$$
$$\left(\prod_{i=2}^{i=n}\left[\left(1-p\right) \cdot f\left(t_i - t_{i-1}\right) + p f_{\lambda_r}\left(t_i - t_{i-1}\right)\right]\right) \cdot$$
$$\left(\left(1-p\right) \cdot S\left(t^* - t_n\right) + p \cdot S_{\lambda_r}\left(t^* - t_n\right)\right)$$

where: $f\left(t_1 - t_0\right) = f_{\lambda_a}\left(t_1 - s^*_{a-1}\right) \cdot \prod_{j=1}^{j=a-1} S_{\lambda_j}\left(u_j\right),$

if t_1 is during period "a",

$$f(t_i - t_{i-1}) = f_{\lambda_a}(t_i - s^*_{a-1}) \cdot S_{\lambda_b}(s^*_b - t_{i-1}) \cdot \prod_{j=b+1}^{j=a-1} S_{\lambda_j}(u_j),$$

if t_i is during period "a" and t_{i-1} is in period "b", and

$$S(t^* - t_n) = S_{\lambda_k}(t^* - s^*_{k-1}) \cdot S_{\lambda_c}(s^*_c - t_n) \cdot \prod_{j=c+1}^{j=k-1} S_{\lambda_j}(u_j),$$

if t^* is in period "k" and t_n in period "c".

When $n = 0$, when no clinical mastitis occurred during lactation, the result is: $L(\lambda_1, \ldots, \lambda_k) = S(t^* - t_0)$.

APPENDIX II

Recurrent formula for probability of count process with Ren parameters

We consider a lactation for which the mastitis Ren hazard λ_r and rate p are constant and for which the Rex hazard is constant by steps over K periods and is equal to λ_k over the kth period $]s^*_{k-1}, s^*_k]$. For any times t_1 and t_2 where $t_1 < t_2$, the number of events occurring between t_1 and t_2 is denoted $N(t_1, t_2)$. For any time t in the interval $]t_0, t^*[$ within period "k", the probability of the event $\{N(t_0, t) = 0\}$ is determined by:

$P(N(t_0,t) = 0) = S_{\lambda_k}(t - t_0)$ for the 1st period $(k = 1)$ and

$$P(N(t_0,t) = 0) = \left[\prod_{j=1}^{j=k-1} S_{\lambda_j}(u_j)\right] \cdot S_{\lambda_k}(t - s^*_{k-1})$$

if $k > 1$.

Let t_w be the time of occurrence of the wth mastitis ($w > 0$), within period "a", such as $t_w < t$, where t is any time within period "k". The probability of the event $\{N(t_w, t) = 0\}$ is determined by a probability mixture for the absence of Ren and Rex mastitis, with a Ren rate p:

$$P(N(t_w,t) = 0) = (1-p) \cdot S_{\lambda_a}(s^*_a - t_w) \cdot$$

$$\left[\prod_{j=a+1}^{j=k-1} S_{\lambda_j}(u_j)\right] \cdot S_{\lambda_k}(t - s^*_{k-1}) + p \cdot S_{\lambda_r}(t - t_w).$$

For any time t within period "k", the probability of the event $\{N(t_0, t) = 1\}$ is determined by:

$$P(N(t_0,t) = 1) =$$

$$\int_{t_0}^{t}\left[\sum_{a \leq k} P(N(t_0, s^*_{a-1}) = 0) \cdot f_{\lambda_a}(t_1 - s^*_{a-1}) \cdot P(N(t_1,t) = 0)\right] \cdot dt_1$$

where $t_1 < t$ denotes the time of occurrence of the first mastitis within an unknown period "a" preceding period "k".

Then for any $w > 1$, the probability of the event $\{N(t_0, t) = w\}$ is determined by the following recurrence formulas:

$$P(N(t_0,t) = w) =$$

$$\int_{t_0}^{t}\left[\sum_{b \leq a \leq k} P(N(t_0, t_{w-1}) = w-1) \cdot f(t_w - t_{w-1}) \cdot P(N(t_w,t) = 0)\right] \cdot dt_w$$

with "a" as the period including time t_w and "b" the period containing time t_{w-1},

$$f(t_w - t_{w-1}) = (1-p) \cdot f_{\lambda_a}(t_w - s^*_{a-1}) \cdot$$

$$\prod_{j=b+1}^{j=a-1} S_{\lambda_j}(u_j) \cdot S_{\lambda_b}(s^*_b - t_{w-1}) + p \cdot f_{\lambda_r}(t_w - t_{w-1}).$$

APPENDIX III

Martingale residual with Ren parameters

When t_i is the date of occurrence of the ith event within period k, with δ_i equalling 1 when there is mastitis and δ_i equalling 0 when there is censorship by drying off, the martingale residuals are defined with the previously introduced notations:

$$r_i = \delta_i - \left[\lambda_k \cdot (t_i - s^*_{k-1}) + \sum_{j=1}^{j=k-1} \lambda_j \cdot u_j\right]$$

if $i = 1$, and

$$r_i = \delta_i - \left[\sum_{t=t_{i-1}}^{t=t_i} \left(\frac{(1-p) \cdot f(t - t_{i-1}) + p \cdot f_{\lambda_r}(t - t_{i-1})}{(1-p) \cdot S(t - t_{i-1}) + p \cdot S_{\lambda_r}(t - t_{i-1})} \right) \right]$$

if $i > 1$, with $f(t - t_{i-1}) =$

$$f_{\lambda_a}\left(t - s^*_{a-1}\right) \cdot S_{\lambda_b}\left(s^*_b - t_{i-1}\right) \cdot \prod_{j=b+1}^{j=a-1} S_{\lambda_j}(u_j),$$

and $S(t - t_{i-1}) = S_{\lambda_a}\left(t - s^*_{a-1}\right) \cdot S_{\lambda_b}\left(s^*_b - t_{i-1}\right) \cdot$

$$\prod_{j=b+1}^{j=a-1} S_{\lambda_j}(u_j), \text{ if } t \text{ is in the period "}a\text{" and}$$

t_{i-1} in the period "b".

REFERENCES

[1] Adkinson R.W., Ingawa K.H., Blouin D.C., Nickerson S.C., Distribution of Clinical Mastitis Among Quarters of the Bovine Udder, J. Dairy Sci. 76 (1993) 3453-3459.

[2] Allore H.G., Erb H.N., Approaches to modeling intramammary infections in dairy cattle, Prev. Vet. Med. 39 (1999) 279-293.

[3] Allore H.G., Jones L.R., Merrill W.G., Oltenacu P.A., A Decision Support System for Evaluating Mastitis Information, J. Dairy Sci. 78 (1995) 1382-1398.

[4] Allore H.G., Erb H.N., Schruben L.W., Oltenacu P.A., A Simulation of Strategies to Lower Bulk Tank Somatic Cell Count Below 500,000 per Milliliter, J. Dairy Sci. 81 (1998) 694-702.

[5] Allore H.G., Schruben L.W., Erb H.N., Oltenacu P.A., Design and Validation of a Dynamic Discrete Event Stochastic Simulation Model of Mastitis Control in Dairy Herds, J. Dairy Sci. 81 (1998) 703-717.

[6] Andersen P.K., Borgan O., Gill R.D., Keiding N., Statistical models based on counting processes, Spinger-Verlag, New York, 1993.

[7] Barnouin J., Geromegnace N., Chassagne M., Dorr N., Sabatier P., Facteurs structurels de variation des niveaux de comptage cellulaire du lait et de fréquence des mammites cliniques dans 560 élevages bovins répartis dans 21 départements français, INRA Prod. Anim. 12 (1999) 39-48.

[8] Bigras-Poulin M., Meek A.H., Martin S.W., McMillan I., Health problems in selected Ontario Holstein cows: frequency of occurrences, time to first diagnosis and associations, Prev. Vet. Med. 10 (1990) 79-89.

[9] Bramley A.J., Dodd F.H., Reviews of the progress of Dairy Science: Mastitis control – progress and prospects, J. Dairy Res. 51 (1984) 481-512.

[10] Clayton D., Cuzick J., Multivariate generalizations of the proportional hazards model, J. R. Stat. Soc. A 148 (1985) Part 2 , 82-117.

[11] Collett D., Modelling Binary Data, Chapman & Hall, Londres, 1991.

[12] Curtis C.R., Erb H.N., Sniffen C.J., Smith R.D., Powers P.A., Smith M.C., White M.E., Hillman R.B., Pearson E.J., Association of parturient hypocalcemia with eight periparturient disorders in Holstein cows, J. Am. Vet. Med. Assoc. 183 (1983) 559-561.

[13] Diggle P.J., Liang K., Zeger S.L., Analysis of longitudinal Data, Clarendon Press, Oxford, 1994.

[14] Donald A., Prevalence estimation using diagnostic tests when there are multiple, correlated disease states in the same animal or farm, Prev. Vet. Med. 15 (1993) 125-145.

[15] Donald A.W., Estimating the prevalence of mastitis and other organ-specific diseases in the presence of within-animal disease correlation and diagnostic test correlation, Prev. Vet. Med. 20 (1994) 113-133.

[16] Donald A.W., Gardner I.A., Wiggins A.D., Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors, Prev. Vet. Med. 19 (1994) 167-187.

[17] Döpfer D., Barkema H.W., Lam T.J.G.M., Schukken Y.H., Gaastra W., Recurrent Clinical Mastitis Caused by *Escherichia coli* in Dairy Cows, J. Dairy Sci. 82 (1999) 80-85.

[18] Fleming T.R., Harrington D.P., Counting Processes and Survival Analysis, John Wiley & Sons, New York, 1991.

[19] Gasqui P., Coulon J.B., Modélisation de l'occurrence des mammites cliniques au sein d'un troupeau: analyse de la surdispersion, Epidémiol. Santé Anim. 32 (1997) 13.21.1-3.

[20] Gasqui P., Pons O., Modèle de prediction de l'occurrence des mammites cliniques chez la vache laitière au cours de ses lactations successives, Epidémiol. Santé Anim. 34 (1998) 145-150.

[21] Goff J.P., Horst R.L., Physiological Changes at Parturition and their Relationship to Metabolic Disorders, J. Dairy Sci. 80 (1997) 1260-1268.

[22] Hortet P., Seegers H., Loss in milk yield and related composition changes resulting from clinical mastitis in dairy cows, Prev. Vet. Med. 37 (1998) 1-20.

[23] Houben E.H.P., Huirne R.B.M., Dijkhuizen A.A., Kristensen A.R., Optimal Replacement of Mastic Cows Determined by a Hierarchic Markov Process, J. Dairy Sci. 77 (1994) 2975-2993.

[24] International Dairy Federation, Recommendations for representation of mastitis-related data,

Bulletin of the IDF, IDF, Bruxelles, 321 (1997) 16-21.

[25] Johnson N.L., Kotz S., Exponential Distribution, Chapter 18, in: Continuous univariate distribution-1, John Wiley, New York, 1970, pp. 207-232.

[26] Johnson N.L., Kotz S., Kemp A.W., Poisson Distribution, Chapter 4, in: Univariate Discrete Distributions, 2nd ed., John Wiley, New York, 1992, pp. 151-198.

[27] Kehrli M.E. Jr, Goff J.P., Periparturient hypocalcemia in cows: effects on peripheral blood neutrophil and lymphocyte function, J. Dairy Sci. 72 (1989) 1188-1196.

[28] Kelton D.F., Lissemore K.D., Martin R.E., Recommendations for recording and calculating the incidence of selected clinical diseases of dairy cattle, J. Dairy Sci. 81 (1998) 2502-2509.

[29] Kossaibati M.A., Hovi M., Esslemont R.J., Incidence of clinical mastitis in dairy herds in England, Vet. Rec. 143 (1998) 649-653.

[30] Lancar R., Méthodes robustes d'analyse de temps de survie multivariés, Rev. Epidémiol. Santé Publique 47 (1999) 287-296.

[31] Lemdani M., Pons O., Estimation and tests in finite mixture models for censored survival data, Statistics 29 (1997) 363-388.

[32] Lescourret F., Coulon J.B., Faye B., Predictive Model of Mastitis Occurrence In the Dairy Cow, J. Dairy Sci. 78 (1995) 2167-2177.

[33] McCullagh P., Nelder J.A., Generalized Linear Models, 2nd ed., Chapman & Hall, London, 1989.

[34] McDermott J.J., Progress in analytic methods – more sophistication or back to basics?, Prev. Vet. Med. 25 (1995) 121-133.

[35] McDermott J.J., Schukken Y.H., A review of methods used to adjust for cluster effects in explanatory epidemiological studies of animal populations, Prev. Vet. Med. 18 (1994) 155-173.

[36] McDermott J.J., Schukken Y.H., Shoukri M.M., Study design and analytic methods for data collected from clusters of animals, Prev. Vet. Med. 18 (1994) 175-191.

[37] McMillan K.L., Duirs G.F., Duganzich D.M., Associations Between Dry Cow Therapy, Clinical Mastitis, and Somatic Cell Count Score with Milk and Fat Production in Ten New Zealand Dairy Herds, J. Dairy Sci. 66 (1983) 259-265.

[38] Morse D., DeLorenzo M.A., Wilcox C.J., Natzke R.P., Bray D.R., Occurrence and Reoccurrence of Clinical Mastitis, J. Dairy Sci. 70 (1987) 2168-2175.

[39] Noordhuizen J.P.T.M., Frankena K., van der Hoofd C.M., Graat E.A.M., Analysis of time at risk (survival) data, Chapter VII, in: Application of quantitative methods in veterinary epidemiology, Wageningen Pers, Wageningen, 1997, pp. 181-200.

[40] Oltenacu P.A., Natzke R.P., Mathematical of the Mastitis Infection Process, J. Dairy Sci. 59 (1975) 515-521.

[41] Rodriguez-Zas S.L., Gianola D., Shook G.E., Bayesian Analysis via Gibbs Sampling of Susceptibility to intramammary Infection in Holstein Cattle, J. Dairy Sci. 81 (1998) 2710-2722.

[42] Schukken Y.H., Casella G., van den Broek J., Overdispersion in clinical mastitis data from dairy herds: a negative binomial approach, Prev. Vet. Med. 10 (1991) 239-245.

[43] Stott A.W., Kennedy J.O., The economics of culling dairy cows with clinical mastitis, Vet. Rec. 133 (1993) 494-499.

[44] Tempelman R.J., Gianola D., Genetic Analysis of Fertility in Dairy Cattle Using Negative Binomial Mixed Models, J. Dairy Sci. 82 (1999) 1834-1847.

[45] Todhunter D.A., Smith K.L., Hogan J.S., Environmental Streptococcal Intramammary Infections of the Bovine Mammary Gland, J. Dairy Sci. 78 (1995) 2366-2374.

[46] Wilson D.J., Gonzalez R.N., Case K.L., Garrison L.L., Gröhn Y.T., Comparison of seven Antibiotic Treatments with No Treatment for Bacteriological Efficacy Against Bovine Mastitis Pathogens, J. Dairy Sci. 82 (1999) 1664-1670.