

Application of survival analysis in breeding for longevity

Natascha Vukasinovic

*Animal Breeding Group, Swiss Federal Institute of Technology
Clausius-Strasse 50, CH-8092 Zurich, Switzerland*

Abstract

Genetic evaluation for the length of productive life (LPL) in livestock species requires specific methods because of following reasons: (i) some animals are still alive at the time of evaluation and their complete LPL is not known, which implies that records of such individuals must be treated as censored; (ii) effects influencing LPL do not act linearly and vary with time; and (iii) the distribution of LPL data is often unknown and extremely skewed. Models based on survival analysis provide appropriate statistical tools for the analysis of LPL. These models are based on a hazard function $\lambda(t)$ which describes the risk of being culled at time t conditional upon survival to time t . The hazard function of an animal is modeled as a product of a baseline hazard function, $\lambda_0(t)$, describing the ageing process, and a vector containing explanatory variables that supposedly influence the culling process and may depend on time. Because the form of $\lambda_0(t)$ is not always known with certainty, $\lambda_0(t)$ can be left unspecified, as in a Cox model. However, with complex models and large amount of data typical for national genetic evaluation, it is computationally advantageous to assume a parametric, e.g., Weibull hazard function. Such a model can also be easily extended to include a random sire or animal effect and enable estimation of genetic parameters for LPL. Routine genetic evaluation for LPL based on a Weibull model has been implemented in many countries worldwide. This paper gives an example of implementation of survival analysis techniques in dairy populations in Switzerland.

1. Introduction

In livestock production, in particular in dairy cows, longevity of breeding stock is a highly desirable trait that considerably affects overall profitability. An increased longevity helps to reduce costs associated with raising or purchasing replacement females. Also, average herd production is increased, due to higher proportion of mature cows in later more productive lactations. An increased longevity enables a greater selection response, because fewer animals have to be replaced and thus, higher selection intensity of females is possible.

From the breeder's perspective, a trait of particular interest is functional longevity, which is independent of production and reflects cow's fertility, health and overall fitness. The importance of functional longevity in dairy breeding has increased during last two decades because of its increasing impact on production costs, and because of growing need to improve functional longevity in many dairy populations heavily selected only on production.

However, considering longevity in breeding programs and genetic evaluation of animals for this trait are generally difficult. The main difficulty is

that a part of the animals is still alive in the moment of genetic evaluation and only the lower bound of their eventual productive life is known. To exclude these records from the evaluation or consider them as exact would lead to biased results. Therefore, such records must be treated as censored. Several researchers tried to circumvent this problem by using indirect longevity indicators such as whether the cow is still alive at certain age (Everett et al., 1976) or lactation (Schaeffer and Burnside, 1974) or by replacing censored records by so-called projected records based on currently available information (Van Raden and Klaaskate, 1993). Unfortunately, the first method is associated with a great loss of information. With the second method, the information available to predict complete length of productive life from partial records is extremely limited (Ducrocq, 1997).

Another difficulty associated with genetic evaluation for longevity is that the overall longevity results from a product rather than from a sum of effects influencing the trait (Beilharz et al., 1993); when at least one of them is defective, the longevity of an individual is impaired (Ducrocq and Sölkner, 1998). Therefore, evaluation methods based on traditional linear models, such as BLUP, although widely used in some countries, cannot properly

account for nonlinearity of longevity data. Also, some of the effects that influence productive life, such as milk production, herd size or management, vary with time.

Furthermore, the distribution of longevity data is extremely skewed and often unknown. Thus, methods based on assumption of normality have only limited use in the analysis of longevity data (Egger-Danner, 1993)

Survival analysis, a statistical method originally developed for research in medicine and engineering, can be used in analyzing longevity data. Survival analysis combines information on dead (uncensored) and alive (censored) individuals, enables a proper statistical treatment of censored records and accounts for nonlinear characteristics of longevity data.

Famula (1981) was the first author who proposed survival analysis as a method to analyze length of productive life in dairy cattle. Smith (1983) and Smith and Quaas (1984) used survival analysis techniques to estimate breeding values of sires based on the length of productive life of their daughters. The techniques were further developed and adjusted for large scale applications by Ducrocq (1994) and Ducrocq and Sölkner (1998). By now, routine genetic evaluation of sires has been implemented in France and research projects are underway in several other European countries.

The aim of this paper is to give insight into basic statistical techniques used in modeling of longevity data and genetic evaluation for longevity in dairy cows. Application of survival analysis in the routine genetic evaluation of sires for the length of productive life will be illustrated using an example of Holstein population in Switzerland.

2. Statistical techniques of survival analysis

Statistical techniques used in survival analysis are aimed at modeling and analysis of response times. The response time of an individual is a positive random variable describing a difference between a given origin point and an end point. In analyzing productive life of dairy cows, the origin point corresponds to the first calving of a cow and the end point ("failure") corresponds to her death or culling. The width of the interval is measured in days, months, or years. Generally, the end point may correspond to any type of event – e.g., recovery from disease or success after insemination - and can be expressed in amount of money spent or earned, kg of milk produced, etc.

A specific feature of survival analysis is that it can accommodate censoring and in this way uses information available on animals that are still alive.

The most frequent type of censoring is right censoring, for which the exact failure time is known to be larger than the observed value. Usual causes of

right censoring are the absence of failure before the end of the study, the loss to follow up or the loss of competing risk. Observation is censored when a cow is still alive at the end of data collection, when she disappears from the herd due to sale to another herd not under testing program, or when the whole herd is withdrawn from the testing program.

Survival analysis can also accommodate left truncated records, for which the origin point lays outside the data collection period. The observation is left truncated when the cow calves for the first time before the data collection starts. Because no information is available for the period prior to begin of data collection, such cow is considered to be at risk only from the beginning of data collection onward. Truncated records are completed records, in contrast to censored records which contain only partial information.

Another type of censoring – left censoring – for which failure occurs prior to a given origin point, is not of interest for applications in animal breeding, because animals that died before the data recording starts are not considered in the analysis.

2.1 Failure time distributions

The analysis of survival data is based on the use of special modeling distributions (Kalbfleisch and Prentice, 1980):

The survivor function $S(t)$:

$$S(t) = \text{Prob}(T \geq t) = 1 - \text{Prob}(T < t) = 1 - F(t)$$

gives the probability that an animal survives at least up to time T . $S(t)$ is the proportion of animals still alive at t . $F(t)$ is the cumulative probability density function.

The density function $f(t)$:

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(t \leq T < t + \Delta t)}{\Delta t} = -\frac{dS(t)}{dt}$$

is a limiting probability that failure will occur between t and Δt .

The hazard function $\lambda(t)$:

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt}$$

is the conditional probability of an individual of dying in the interval $(t, t+\Delta t]$, given it was alive at time t .

All these function are interrelated and each of them can be derived from another.

2.2. Empirical estimation of the survivor function

Empirical survivor function gives an idea about the distribution of survival times. The empirical

survivor function can be computed using a Kaplan-Meier's formula (Kaplan and Meier, 1958):

$$\hat{S}_{KM}(t) = \prod_{k|T_{[k]} < t} \left(\frac{n_k - d_k}{n_k} \right)$$

where $\hat{S}_{KM}(t)$ is the value of the survivor function at time t , $T_{[k]}$ represents the ordered failure times, from the first occurrence of the failure to the last one, n_k is the number of animals at risk at $T_{[k]}$, and d_k is the number of animals that actually died at $T_{[k]}$. $\hat{S}_{KM}(t)$ is called the product-limit or Kaplan-Meier estimate of the survivor function. $\hat{S}_{KM}(t)$ is also the generalized maximum likelihood estimate of $S(t)$ over all possible distributions (Kaplan & Meier, 1958).

From $\hat{S}_{KM}(t)$ the empirical hazard function $\lambda(t)$ can be computed.

Empirical estimates of S and λ are useful in preliminary studies: they allow to reveal the general features of the underlying distributions of interest, they facilitate the choice of models for further analyses, and enable to check the validity of these models (Ducrocq, 1992).

2.3. Regression models

In many cases, the exact characteristics of the survivor function or the density function are not known, but some information is available on how the failure rate $\lambda(t)$ changes over time. Therefore, models for survival analysis are usually built from a hazard function, which measures the risk of failure of an individual at time t . A general assumption is that a hazard function for each animal has a common basic form, called baseline hazard function, which is equal for all animals and represents, in some sense, the overall mean. The baseline hazard function of an individual is modified by the influence of the effects that supposedly influence the culling process, called 'stress factors'.

A survival model in its general form can be written as

$$\lambda(t, \mathbf{x}_i) = \lambda_0(t) \exp(\mathbf{x}'\boldsymbol{\beta})$$

where $\lambda(t, \mathbf{x}_i)$ is the hazard for animal i . $\lambda_0(t)$ is the baseline hazard function, describing the aging process of an animal and depending only on time. In parametric models, the baseline hazard function is assumed to follow a certain exponential distribution, usually an exponential (exponential model) or Weibull distribution (Weibull model). The baseline hazard function can also be left completely arbitrary, as in semiparametric or Cox model. \mathbf{x}_i is the vector of explanatory variables (stress factors) that supposedly

influence the culling process, and $\boldsymbol{\beta}$ is the vector of regression coefficients.

2.4. Estimation procedures

Estimation of effects of particular covariates on survivor curve is usually based on maximum likelihood methods. The construction of the likelihood function, however, requires a different approach for parametric and semiparametric models.

For parametric models, it could be shown that the contribution to the likelihood for an uncensored observation equals the value of the density function at failure time. The contribution to the likelihood function for a censored observation equals the value of the survivor function at censoring time (Ducrocq, 1987). The full likelihood function is then obtained by multiplying conditionally independent contributions for all records. The estimates of the parameters are obtained by maximizing this function or its logarithm (Ducrocq, 1997). The parameters of the baseline hazard function are estimated jointly with the regression coefficients.

For semiparametric models (Cox model) a different approach must be used. The estimation of $\boldsymbol{\beta}$ can be performed by maximizing a likelihood function without making any assumptions about $\lambda_0(t)$. Cox (1972) developed a concept of partial likelihood. A partial likelihood function is a part of the full likelihood function that does not depend on $\lambda_0(t)$. It can be obtained as the marginal likelihood of the ranks of failure times.

After maximizing the partial likelihood function, the resulting estimates of $\boldsymbol{\beta}$ can be used to compute baseline hazard function or the baseline survivor function, using an approach similar to that used in computation of the Kaplan-Meier estimator.

2.5. Proportional hazard model and its generalizations

One of the most frequently used models in analysis of productive life is a proportional hazard model, for which the hazard of two animals A and B remains proportional over the entire period of time with a constant hazard ratio $\exp(\mathbf{x}_A - \mathbf{x}_B)' \boldsymbol{\beta}$. In most cases, however, the assumption of proportional hazard does not hold, because the hazard for individual cows vary differently over time, depending on current culling criteria.

A possible alternative that retains the simplicity of the proportional hazard model is known as stratification. Stratification is characterized by definition of a different baseline hazard function for each stratum – a subclass of individuals with, e.g., same year of birth, herd, or production level.

Another powerful generalization of the proportional hazard model is the induction of time-dependent explanatory variables that account for variation in culling policy over time, such as herd by year or lactation number. In this way, the time axis is partitioned into several intervals. The assumption of proportional hazard holds within each interval, but the hazard ratio changes from one interval to another. With time-dependent covariates, several levels of a fixed effect can be associated with a single record of productive life.

2.6. Cox vs. Weibull model in animal breeding

Cox model has been repeatedly used in animal breeding because it does not require knowledge about the distribution of survival times which is, in many cases, unknown. However, when a Cox model contains several time dependent covariates with a large number of levels, the estimation of coefficients in β may lead to extremely tedious computations. Such a model cannot be applied to usually very large data sets used in a national sire evaluation.

But, when the baseline hazard function has a known parametric form, the estimation of β and $\lambda_0(t)$ is generally easier (Cox and Oakes, 1984). Intuitively, the most appropriate parametric form of the baseline hazard function is a Weibull distribution. In a Weibull regression model, the baseline hazard function, $\lambda_0(t) = \lambda\rho(\lambda t)^{\rho-1}$ is described by two parameters, λ and ρ , which define the scale and the shape of the baseline hazard function.

The choice of the Weibull distribution results from the simplicity of the Weibull survivor function $S_0(t) = \exp(-(\lambda t)^\rho)$ combined with its flexibility: a Weibull regression can model constant ($\rho = 1$), increasing ($\rho > 1$), and decreasing ($\rho < 1$) hazard. If an approximation of $\lambda_0(t)$ is possible, further analyses can greatly be facilitated. An additional advantage of a Weibull model is its easy extension to mixed survival models that can include correlated random effects such as relationships among sires (Ducrocq and Casella, 1996).

2.7. Frailty models

Proportional hazard models, either parametric or semiparametric, can be extended to include random (e.g., genetic) effects. Such mixed survival models are referred to as frailty models. The frailty term v_m describes the genetic effect which multiplicatively influences the hazard of each individual or a group of individuals. For example, hazard for cow m can be written as

$$\lambda(t; \mathbf{x}_m) = \lambda_0(t) v_m \exp(\mathbf{x}_m' \beta).$$

The frailty component explains a portion of the variance that cannot be accounted for by the model containing only fixed effects. A simple transformation $s_m = \log v_m$ allows the inclusion of the frailty term in the term $\exp(\mathbf{x}'\beta)$ of the usual regression model. If \mathbf{z} represents the incidence vector for the random effect $\mathbf{s} = \{s_m\}$ the mixed survival model can be written as

$$\lambda(t; \mathbf{x}, \mathbf{z}) = \lambda_0(t) \exp\{\mathbf{x}'\beta + \mathbf{z}'\mathbf{s}\}$$

Traditionally, a gamma distribution has been assumed for the frailty term \mathbf{v} because of its flexibility and mathematical convenience (Ducrocq et al., 1988; Aslanidou and Dey, 1996). Unfortunately, a gamma distribution do not have the theoretical appeal of the normal distribution usually used in animal breeding under the assumption of the infinitesimal polygenic model (Ducrocq and Casella, 1996). However, it can be shown that the estimates obtained for the parameters of the gamma distribution of \mathbf{v} are relatively large in dairy cattle populations. This means that \mathbf{v} had an approximate log-normal distribution, i.e., \mathbf{s} is approximately normally distributed (Ducrocq et al., 1988). Therefore, it has been suggested to account for genetic relationships among animals by assuming a multivariate normal distribution for \mathbf{s} (Ducrocq, 1997).

A number of approaches has been proposed to estimate parameters of the distribution of \mathbf{v} - so-called hyperparameters. For a Cox model, Klein (1992) suggested the use of an EM algorithm, with iterative estimation of \mathbf{v} , fixed effects β and the baseline hazard function, followed by the estimation of the distribution of frailty term given its estimates $\hat{\mathbf{v}}$. For a Weibull model, if a gamma distribution of the frailty term is assumed, hyperparameters of its distribution can be obtained by exact algebraic integration of \mathbf{v} from the likelihood function (Follman and Goldberg, 1988). This approach has been applied in Bayesian context by Ducrocq and Casella (1996). To estimate \mathbf{v} along with the hyperparameters of its distribution, Monte Carlo techniques have been suggested (e.g., Korsgaard, 1996), but their use on large data sets and with complex models can computationally be very demanding. Ducrocq and Casella (1996) proposed a Bayesian approach using a Laplacian approximation of the marginal posterior density of the hyperparameters of the distribution of \mathbf{v} or \mathbf{s} . This approach is applicable to both Cox and Weibull model. It has been proved very efficient for large applications.

2.7.1. Definition of heritability

A distribution attached to the elements of v_q or s_q , when v_q denotes the genetic, e.g. sire effect, and $s_q = \log v_q$, naturally leads to a definition of heritability. The heritability of a survival trait on the logarithmic scale is

$$h_{\log}^2 = \frac{4\text{Var}(s)}{\text{Var}(\log \text{LPL})} = \frac{4\text{Var}(s)}{\text{Var}(s) + \frac{\pi^2}{6}}$$

$\pi^2/6$ is the variance of the extreme value distribution corresponding to the error variance in survival models.

Heritability on the original scale can be approximated using a Taylor series expansion of $\log(\text{LPL})$ around its mean (Ducrocq, 1997). The heritability on the original scale amounts to 0.15 - 0.20 (Ducrocq and Sölkner, 1998b). These values are usually higher than on the logarithmic scale and, in particular, considerably higher than the heritability obtained from linear models that rarely exceeds a value of 0.10 (e.g., Everett et al, 1976). This indicates that the survival analysis really provides a better description of longevity data and a more complete extraction of the genetic variance. Consequently, a faster genetic progress can be expected if longevity is included in selection programs.

3. Example: Genetic evaluation for the length of productive life in Swiss Holsteins

3.1 Material and methods

Analysis of length of productive life (LPL) and estimation of breeding values was conducted using data provided by the Swiss Holstein Breeders' Association. The data included all cows calved from April, 1, 1980 onward. LPL was defined as the number of days between the first calving and the last test day entered in the databank. Records on cows with last test day after February 1, 1998, changing herds during their herd life or disappearing from the data file were considered as censored. For cows with first calving before April 1, 1980, only the part of LPL corresponding to lactations started after that date was considered (left-truncated records). Breeding values were estimated for all sires having at least 6 daughters in the data. Also, maternal grandsires (MGS) of the cows were included. The pedigree file comprised all sires and two generations of their male ancestors.

Characteristics of Holstein data used in genetic evaluation are given in Table 1.

Table 1: Characteristics of Holstein data used for genetic evaluation

Item	
Total records (no)	224 847
Censored (%)	19
Left truncated (%)	5
Sires with daughters (no)	1 656
MGS (no)	1 455
Total sires in pedigree (no)	3 234
¹ Sire variance (σ_s^2)	0.036
h_{\log}^2	0.072
h_{orig}^2	0.184

¹ Sire variance was estimated in a separate analysis using a subset of the data

The following Weibull mixed survival model was used:

$$\lambda(t) = \lambda_0(t) \exp \{ \text{hys}_i(t') + \text{fc}_j + \text{pa}_k(t_1) + \text{ls}_i(t_1, t_2) + \text{m}_m(t_1) + \text{p}_n(t_1) + \text{s}_o \}$$

where $\lambda(t)$ is the hazard of the cow, t days after her first calving and:

- $\lambda_0(t)$ is the baseline hazard function assumed to follow a Weibull distribution with $\rho = 1.53$;
- $\text{hys}_i(t')$ is the random time-dependent herd-year-season effect at calendar date t' with changes on March 1 and October 1 each year. The distribution of the effect $\mathbf{hys} = \{\text{hys}_i\}$ was assumed to be log-gamma (γ, γ), with $\gamma = 3.75$. This effect is considered to account for yearly and seasonal changes in herd management and different culling policies in individual herds;
- fc_j is the fixed time-independent effect of age at first calving with 7 classes. This effect is included because cows that first calved at older age are expected to have a shorter productive life;
- $\text{pa}_k(t_1)$ is the fixed time-dependent effect of parity. The effect of parity is considered to account for increased risk of culling younger cows. Parities 1, 2, 3, 4, and 5 were considered;
- $\text{ls}_i(t_1, t_2)$ is the fixed time-dependent combined effect of parity, t_1 days after the first calving, and of stage of lactation, t_2 days after the current calving with changes at $t_2 = 0, 30, 60, 180, 240, \text{ and } 300$. This effect is included to account for changing intensity of culling during each lactation;
- $\text{m}_m(t_1)$ and $\text{p}_n(t_1)$ are the time-dependent effects of within-herd deviations for milk yield (5 classes) and sum of fat and protein content (5 classes) based on age-adjusted 305d lactation records, with changes at each new calving date. Uncensored records on first lactating cows with no lactation record are arbitrarily assigned to class 2 for milk yield and class 3 (average) for fat and protein content. The

effects of within-herd milk yield and fat and protein content are included to account for culling because of poor milk production;

$-s_o$ is the random effect of the sire of the cow. Sire effects $\mathbf{s} = \{s_o\}$ were assumed to follow a multivariate normal distribution with variance-covariance matrix $\mathbf{A} \sigma_s^2$.

All computations have been conducted using the software 'The Survival Kit 3.0' (Ducrocq and Sölkner, 1998a). After the sire variance had been estimated on a subset of data, the obtained value of σ_s^2 has been used in a subsequent genetic evaluation.

3.2. Results for fixed effects

To facilitate the interpretation, the results are expressed as relative culling rate, or relative risk of being culled, defined as the ratio between the estimated risk of being culled under the influence of a certain environmental or genetic effect, $[\exp(\hat{\beta})]$, and the average risk, which is usually set to one, $[\exp(0)]$. In the model for genetic evaluation of Holstein cows following effects have been set to one: 4th group of age at first calving, 4th parity, lactation stage from 181 to 270 days within each parity, and the 3rd (average) within-herd class of milk yield and fat and protein content within each lactation.

Figure 1 shows the relative culling rates for different ages at first calving. Although the differences among age groups are not large, a slight increase in culling risk can be observed for cows with very old (> 37 mo) age at first calving.

Figure 2 shows the estimates for the effect of parity. Compared with the average risk (4th lactation),

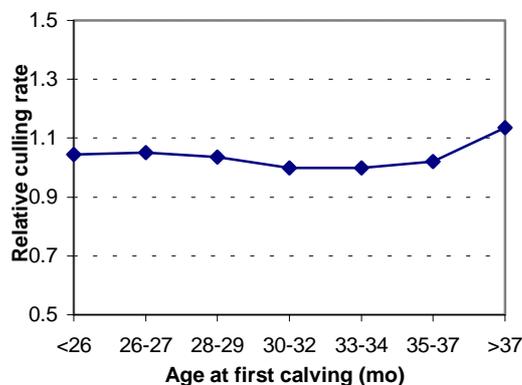


Figure 1: Estimates of the effect of age at first calving

first-lactating cows are at considerably higher risk of being culled. This can be explained through changing intensity of selection, which is higher in the first than in later lactations. Cows after 4th lactation also have increased relative culling rates.

The estimates for the effect parity x stage of lactation for parities 1, 2, and 3 are shown in Figure 3. Relative culling rates increase linearly from the beginning to the end of lactation, reaching the maximum 300 days after calving. This trend is in accordance with the changing selection pressure during the lactation, and more intensive culling of non-pregnant cows during the dry period. Only in the first lactation, an increased risk of culling can be observed during the first 30 days of lactation, indicating intensive selection of heifers in the beginning of lactation.

Figure 4 shows the influence of within-herd deviation of milk yield on the relative culling rates. Within each lactation, cows producing less than 80% of the herd average are at 3-4 x higher risk than their herdmates with average production. High producing cows are less likely to be culled. These results illustrate the great influence of voluntary culling for low production.

The influence of within-herd deviation of fat and protein content can be seen in Figure 5. Except for the first lactation, in which the culling rates are mostly influenced by grouping animals without a lactation record, cows with a high fat and protein content are at slightly lower risk than those with fat and protein content under the herd average. Differences among the classes are smaller than for milk yield.

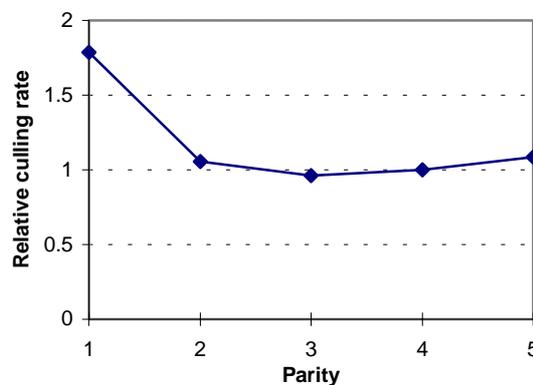


Figure 2: Estimates of the effect of parity

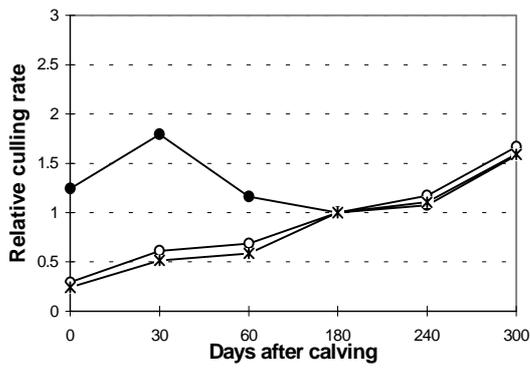


Figure 3: Estimates of the effect of stage of lactation for lactations 1 (●), 2 (○) and 3 (*)

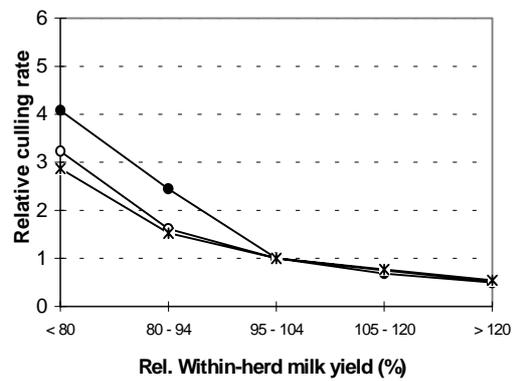


Figure 4: Estimates of the effect of relative within-herd milk yield for lactations 1 (●), 2 (○) and 3 (*)

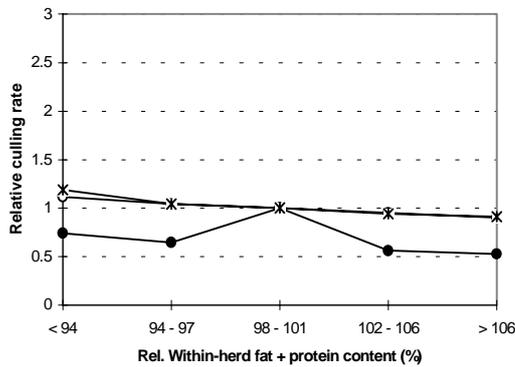


Figure 5: Estimates of the effect of relative within-herd fat and protein content for lactations 1 (●), 2 (○) and 3 (*)

3.3. Sire breeding values

The estimated sire effects (\hat{s}) for 3 234 sires ranged from -0.57 to 0.44 , corresponding to relative culling rates between 0.55 and 1.55 . Negative values of \hat{s} indicate lower risk of culling and, thus, increased LPL of daughters. To enable easier understanding of breeding values for LPL by the farmers, sire breeding values can be expressed in genetic standard deviation units, computed as $\hat{s}/\sqrt{\sigma_s^2}$ after changing the sign. One genetic standard deviation unit corresponds to roughly 150 days of productive life.

Given the sire effect \hat{s} , it is possible to compute an expected survivor curve for daughters of each sire. Figure 6 presents such a curve for three sires with estimated breeding values of $+1$, 0 , and -1 genetic standard deviation units, assuming average levels of the effects of herd-year-season, age at first calving, and within-herd deviations for milk yield and fat and protein content. The variations of culling rate during each lactation are clearly indicated.

The theoretical curves can be used to compute the fraction of daughters still alive at any time after first calving. Cumulating these fractions over the whole

range of productive life, a more intuitive indicator of sire's genetic value can be computed, namely, an expected LPL of daughters in days. Breeders' associations in Switzerland have chosen this parameter to be published in sire catalogs. Sire breeding values are expressed in deviations from a fixed base of 1 018 days of LPL, which corresponds to the average expected LPL of daughters of all Holstein sires born between 1981 and 1985.

The reliability of sires' estimated breeding values can be computed using the common formula:

$$R^2 = n / (n + (4-h^2)/h^2)$$

where R^2 is the reliability of the estimated breeding value, n is the number of uncensored daughters, and h^2 is heritability of LPL (0.184 for Holstein). The reliability can be increased through considering relationships among sires.

The reliability of breeding values for LPL depends on the number of uncensored daughters and not on the total number of daughters. Therefore, sires transmitting shorter LPL will be known with certainty earlier than other sires, because a greater proportion of their daughters will be culled.

4. Conclusions and perspectives

Survival analysis is an appropriate method for analysis of longevity data. Survival analysis can be considered to be the method of choice for genetic evaluation of sires for LPL at national and international level. With rapid development of powerful computers and appropriate software, it is expected that in the future genetic evaluation will be based on animal model rather than current sire or sire-MGS model.

However, to make full use of survival analysis techniques in animal breeding, further research is needed. Future efforts should be focused on improvement of accuracy of the results for young individuals, obtained from data with an increased proportion of censored records.

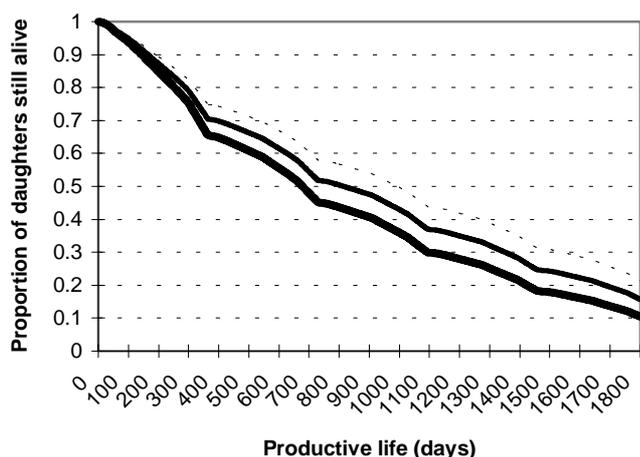


Figure 6: Expected survivor curve of sires with breeding values of +1 (dashed line), 0 (solid line), and -1 (thick line) genetic standard deviation units

In this context, inclusion of some auxiliary traits, such as type traits, should be considered. This implies proper estimation of genetic covariances between these traits and LPL, which might indicate that multivariate survival analysis techniques are needed. Survival models in the future should be extended to other genetic model, e.g. to include information on major genes or genetic markers.

References

Aslanidou, H., Dey, D. K. (1996): Model determination for multivariate data under gamma frailty models. Proceedings of the Biometrics Section, American Statistical Association, Chicago: 292-297.

Beilharz, R.G., Luxford, B.G., Wilkinson, J.I. (1993): Quantitative genetics and evolution: Is our understanding of genetics sufficient to explain evolution? *J. Anim. Breed. Genet.* 110: 161-170.

Cox, D. R. (1972): Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B.* 34: 187-220.

Cox, D. R., Oakes, D. (1984): *Analysis of survival data*. Chapman and Hall, London, England.

Ducrocq, V., Quaas, R.L., Pollak, E.J., Casella, G. (1988): Length of productive life in dairy cows: 2. Variance component estimation and sire evaluation. *J. Dairy Sci.*, 71: 3071-3079.

Ducrocq, V. (1987): An analysis of length of productive life in dairy cattle. Ph.D. Diss., Cornell Univ., Ithaca, NY.

Ducrocq, V. (1992): Survival analysis. Workshop on advanced biometrical methods in animal breeding. Flawil, Switzerland.

Ducrocq, V. (1994): Statistical analysis of length of productive life for dairy cows of the Normande breed. *J. Dairy Sci.* 77: 855-866.

Ducrocq, V. (1997): Survival analysis, a statistical tool for longevity data. 48th Annual Meeting of the EAAP, Vienna, Austria.

Ducrocq, V., Casella, G. (1996): A Bayesian analysis of mixed survival models. *Genetics Selection Evolution* 28: 505-529.

Ducrocq, V., Sölkner, J. (1998a): 'The Survival Kit V3.0' - a package for large analyses of survival data. Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia, Vol. 27: 447-448.

Ducrocq, V., Sölkner, J. (1998b): Implementation of a routine breeding value evaluation for longevity of dairy cows using survival analysis techniques. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia, Vol. 23: 359-362.

Egger-Danner, C. (1993): Zuchtwertschätzung für Merkmale der Langlebigkeit beim Rind mit Methoden der Lebensdaueranalyse. Ph.D. Diss., Univ. für Bodenkultur, Vienna, Austria.

Everett, R. W., J. F. Keown, and E. E. Clapp (1976): Production and stayability trends in dairy cattle. *J. Dairy Sci.* 59:1532-1539.

Famula, T.R. (1981): Exponential stayability model with censoring and covariates. *J. Dairy Sci.* 64: 538.

Follmann, D. A., Goldberg, M. S. (1988): Distinguishing heterogeneity from decreasing hazard rates. *Technometrics* 30: 389-396.

Kalbfleisch, J.D. and Prentice, R.L. 1980. *The statistical analysis of failure time data*. Wiley, New York, NY.

Kaplan, E.L., Meier, P. (1958): Nonparametric estimation from incomplete observations. *J. Amer. Stat. Ass.*, 53: 457-481.

Klein, J.P. (1992): Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, 48: 795-806.

Korsgaard, I.R. (1996): Genetic analysis of survival data - a Gibbs sampling approach. International Workshop on genetic improvement of functional traits in cattle, Gembloux, Belgium.

Schaeffer, L. R., and E. B. Burnside (1974): Survival rates of tested daughters of sires in artificial insemination. *J. Dairy Sci.* 57:1394-1400.

Smith, S.P. (1983): The extension of failure time analysis to problems of animal breeding. Ph.D. Diss., Cornell Univ., Ithaca, NY.

Smith, S.P., Quaas, R.L. (1984): Productive lifespan of bull progeny groups: failure time analysis. *J. Dairy Sci.* 67: 2999-3007.

Van Raden, P. M., and E. H. J. Klaaskate (1993): Genetic evaluation of length of productive life including predicted longevity of live cows. *J. Dairy Sci.* 76:2758-2764.