



HAL
open science

Survival analysis of longevity in dairy cattle on a lactation basis.

A. Roxstrom, Vincent Ducrocq, E. Strandberg

► **To cite this version:**

A. Roxstrom, Vincent Ducrocq, E. Strandberg. Survival analysis of longevity in dairy cattle on a lactation basis.. *Genetics Selection Evolution*, 2003, 35 (3), pp.305-318. hal-02669386

HAL Id: hal-02669386

<https://hal.inrae.fr/hal-02669386v1>

Submitted on 31 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Survival analysis of longevity in dairy cattle on a lactation basis

Anki ROXSTRÖM^{a*}, Vincent DUCROCQ^b,
Erling STRANDBERG^a

^a Department of Animal Breeding and Genetics, PO Box 7023,
Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden

^b Station de génétique quantitative et appliquée,
Institut national de la recherche agronomique, 78352 Jouy-en-Josas Cedex, France

(Received 4 July 2001; accepted 5 November 2002)

Abstract – An analysis of longevity in dairy cattle on a lactation basis is proposed. The approach allowed each lactation to have its own baseline hazard function, which gives a better description of the hazard than traditional analyses of the whole length of life. As a consequence, the overall fit of the model to the data was improved and fewer time-dependent variables were needed. Longevity on a lactation basis was defined from one calving to the next instead of from the first calving to culling. However, no new information was added and it was still the overall risk of being culled that was modelled. It is shown that no cow effect is needed in the lactation basis model because a censored record is not complete, a cow can appear as uncensored only once, and a cow cannot be censored after having been culled. Different subdivisions of the stage of lactation effect were tested and the first ten days of lactation were shown to correspond to an increased risk of being culled. There were no major differences in sire variance between the longevity analysed on a lactation basis and longevity based on the entire length of life.

dairy cows / culling / frailty models / failure time analysis / heritability

1. INTRODUCTION

In many of the current genetic evaluation models for longevity (the length of productive life) in dairy cattle that are based on survival analysis, a lactation by stage of lactation (*lactsta*) effect is included as the time-dependent covariate (in France, Germany, the Netherlands, Denmark, and Italy – see for example Interbull Bulletin N° 21, and [5,6,20]). The value of such covariates varies with time and has proven to lead to a more precise modelling of changes in culling policies over time [4,6]. The stage of the lactation effect is included to account for changes in the culling risk within lactation. The sole baseline hazard function cannot account for these changes, in particular because after

* Correspondence and reprints
E-mail: Anki.Roth@svenskmjolk.se

a same number of days of productive life, cows may have reached different lactation numbers or lactation stages. In general, culling is more intense late in lactation, when production is lower, when it is known whether the cow is pregnant or not, and when her carcass value is higher [2, 5, 13]. Quite naturally, there is an increased risk of culling with an advancing age. There are also differences in the distribution of culling reasons between different lactations, and the time for culling differs between lactations, *e.g.* in later lactations, more cows are culled early in lactation [2].

The changes in classes of the *lactsta* factor are defined somewhat empirically in order to model changes by a piecewise constant function. The interpretation of the *lactsta* factor must be done after combining it with the baseline hazard function. From such graphs it is tempting to conclude that the hazard pattern is more or less the same for each lactation, with a regular increase of the risk during the lactation, and that the regularity is partly broken by the arbitrary choice of changes for the *lactsta* factor [5, 7]. A more regular curve would imply a more precise stage definition, *i.e.* more classes of stages in lactation, but it would add considerably to the computational demand. An alternative is to model the hazard for each lactation separately. The hypothesis is that a model in which survival is looked at on a lactation basis rather than on the entire length of productive life would result in a better description of the *lactsta* factor, between as well as within lactations. This better description may mean that the stage of the lactation effect is completely included in the baseline and that therefore it is not needed in the model, or at least, that the number of stages could be reduced [7].

Longevity is often measured by the use of the date of the first calving as the starting point and the date of culling as the end-point [5, 20]. For the alternative model, each lactation of a cow is treated as one survival measure: the origin point is the date of calving, the end-point is the date of culling or the date of the next calving (whichever comes first), and the record is censored if the cow has a next calving. Cows with more than one lactation will have repeat records and therefore one would expect inclusion of a cow effect. However, this paper will show that the approaches of likelihood construction are equivalent for the two models, and that a cow effect is not needed in the lactation basis model. It should be stressed that no new information is added when analysing longevity on a lactation basis. The trait of interest is in both cases the overall risk of failure. Applications of survival analysis on a lactation basis can be found in epidemiology [2, 12, 13].

The expected attractive features of the lactation basis approach is that it should allow a better fit of the model to the data – through a smoother description of changes in risk along lactation – and an easier handling of the data: since it is no longer necessary to consider the whole herdlife of a cow globally, lactations can be processed separately, as it is often the case in national databases [7].

Also, most effects in survival analyses models are defined within a lactation and are handled as such. The main purpose of this study was to compare the analysis of longevity on a lactation basis *vs.* the analysis on the entire length of life, with respect to the fit of the model to the data and certain parameters.

2. MATERIALS AND METHODS

2.1. Data and definition of censoring

This study was based on survival information collected from altogether 733 492 cows from the Swedish milk recording scheme, which comprises 80–85% of all cows. Cows born between 1986 and 1996 of the Swedish Red and White (SRB) breed were included. The longevity measure t was defined in two ways: time from the first calving to culling or the end of data (PL), and on a lactation basis as time from calving to next calving or culling (LACTPL). The information on the length of the productive life consisted of calving and culling dates. For the actual time period, the overall culling rate in Sweden for SRB cows averaged 38% for all cows and 28% for first lactation cows [17].

The records of cows were excluded if they had: missing sire identification; missing or erroneous herd identification; age at first calving outside 18–42 months; or incorrect calving or birth dates (altogether 145 040 cows). The records of cows changing herds (9504 cows), and the records from herds with fewer than five uncensored observations for productive life over the whole period (7446 cows) were also excluded. Only records from sires with 50 daughters or more were kept (8720 cows were excluded). The sire effects were considered as normally distributed random effects, since there has not been a strong selection for longevity, something that might violate this assumption. Cows in herds with uneven culling frequency (defined as herds having fewer than five cullings in one year-season (ys), no cullings in the ys before, and more than ten cullings in the ys before that) were excluded from the data set (1002 cows). Such herds were assumed to report culling dates incorrectly. A total of 2766 sires and 9210 herds were included in the analyses.

The following categories of records were considered as censored:

- cows still alive at the end of the study period;
- cows in herds disappearing from the recording system. Such herds had no first calvings in the last year, 1996, and the cows were censored on 31st December of the last year with any first calving in that herd (28 808 cows);
- cows sold to a new herd were censored on the last calving date (2983 cows). Such cows may not be treated as any other cow in the purchased herd. Excluding them from the analysis in their second herd is a precaution to avoid potential biases;

- cows with calving intervals, or interval between last calving and the end of data, longer than 1.5 years were censored at the last calving date (18 764 cows). Such cows were assumed to have had a new calving or have been culled, but the dates had not been reported. Cows censored at the first calving date were excluded (10 034 cows), because they did not contribute any information.

Since the zero values for the length of productive life cannot be handled, cows culled or censored at the calving date (11 013 cows) were included with the length of a productive life equal to one day for LACTPL. After these rules were applied, about 37% and 72% of the records were (right-) censored, for PL and LACTPL, respectively. In total 538 783 cows were analysed. The average length of lactation was 344 and 239 days for censored and uncensored cows, respectively. The average length of productive life (PL) of uncensored records was 594 days.

2.2. Models

One preliminary LACTPL analysis was done with a Cox model [3] to: (1) check the validity of the Weibull model; (2) decide how to handle the effect of the stage of lactation in the lactation basis model; and (3) decide if stratification by lactation should be implemented. The model used in the Cox analysis was the same as described below for Weibull, except that the baseline was left arbitrary and that the effect of the stage of lactation was not included because it cannot be estimated in the Cox model. One baseline hazard function, $h_{0,s}(t)$, was defined for each lactation (subscript 0 designates a baseline hazard, subscript s relates to stratum s). A careful examination of the baselines in a stratified Cox model is a way to check the proportional hazards assumption. From each hazard function it is possible to semi-parametrically estimate a baseline survivor function $S_{0,s}$, and if a Weibull model is adequate, a plot of $\ln(-\ln \hat{S}_{0,s})$ versus $\ln t$ should give a straight line with a slope equal to ρ , one of the two Weibull parameters [14]. Furthermore, if these lines computed for different strata run parallel, the proportional hazard assumption holds and a common baseline can be used over the strata. If there are segments that are straight but with different slopes from one segment to the other, it may be concluded that the Weibull assumption holds within the segment, but that a change in the Weibull parameters is needed from one segment to the next, which is what is done when a stage of lactation effect is included.

When the Weibull model was used, the intercept ($\rho \ln \lambda$), ρ and the sire variance, σ_s^2 were estimated simultaneously with the other fixed effects in the model. The separate estimates of the random herd-year-season effects were not of interest *per se*. They were considered here as nuisance parameters. Therefore the herd-year-season effect was algebraically integrated out according to [8]

and only its variance was estimated. For the Cox model, the herd-year-season effect cannot be algebraically integrated out. It was therefore estimated but its variance was assumed to be known from a Weibull analysis. Also, to maintain computational feasibility and with virtually no consequences on the results, the sire variance was fixed at the value obtained from a Weibull analysis.

The only difference in the statistical models between PL and LACTPL was the number of the baseline hazard functions. For PL, one baseline hazard function $h_0(t)$ was defined. For LACTPL, one, two or six different baseline hazard functions $h_{0,s}(t)$ were defined, corresponding to one common baseline for all lactations, first *versus* later lactations, or one baseline per lactation, respectively. In the case of LACTPL there were repeated records for each cow and therefore one would expect the inclusion of a cow effect as well. In Appendix it is shown that this is not necessary. The construction of the likelihood is completely equivalent for the two models. The reasons are that a censored record is not complete, that a cow can appear as uncensored only once, and that a cow cannot be censored after having been uncensored [18]. This result can also be interpreted as the mere consequence of another presentation of the same information, without any modification of its content: knowing that an animal was culled during its 4th lactation is equivalent to considering that it survived until the end of its first lactation, then survived until the end of its second lactation given it was alive at the end of the 1st and so on, until culling was observed during the 4th lactation.

The model applied to the individual m was:

$$h(t; \mathbf{x}_m, \mathbf{z}_m) = h_0(t) \exp\{\mathbf{x}'_m(t)\beta + \mathbf{z}'_m u\}$$

where:

$h_0(t)$ is the baseline hazard; unspecified in the Cox analysis and assumed to follow a Weibull hazard distribution ($h_0(t) = \lambda\rho(\lambda)t^{\rho-1} = \rho t^{\rho-1} e^{\rho \ln \lambda}$) with parameters λ and ρ in the Weibull analyses, and t is the time in days from the first calving for PL and days from the previous calving for LACTPL;

β contains the possibly time-dependent covariates affecting the hazard, with $\mathbf{x}'_m(t)$ being the corresponding design vector, and

u is a vector of random variables with associated incidence vector \mathbf{z}'_m .

The independent covariates included in the model are as follows:

hys

The random time-dependent effect of the herd-year-season class; where the years were from 1988 to 1996, seasons were January–July and August–December. The *hys*-effect was assumed to follow a log-gamma distribution

with (without loss of generality, [11]) both parameters equal to γ . This effect was algebraically integrated out in the analysis. This was not possible to do with the Cox model [8] at least with the software used, and therefore $\gamma = 4.9$ (estimated from a Weibull analysis) was used.

lact-ys

Fixed the time-dependent effect of the lactation-year-season, the lactations were from 1 to 6⁺, same season classes as above.

lact-age

Fixed the time-dependent effect of lactation-age at the first calving, classified as 18–20, 21, 22, . . . , 39, 40–42 months (21 classes), the lactations were from 1 to 6⁺.

lact-stage-peak

Fixed the time-dependent effect of the lactation number – stage of lactation – peak yield of the lactation class. The lactations were from 1 to 6⁺, stages changed twice (at day 0, 11 or 0, 151), or three times (at day 0, 11, 151 or 0, 11, 306) per lactation. The peak test-day yield in a given lactation was expressed as the deviation from her herdmates in that herd-year. Standardised deviations were calculated using the herd-year mean (m_{hy}) and the overall phenotypic standard deviation of test-day peak yield within herd (SD_{hy} , averaged over all herds). The m_{hy} and SD_{hy} were calculated for the first lactation and all later lactations separately. The cow's actual peak yield in each lactation was used, but in lactations after the first it was deviated from m_{hy} of all later lactations. These standardised deviations were used to place the cow in one of 13 classes, the cut-off points chosen such that the classes were expected to contain (starting from the lowest yields) 4 classes with 2.5% each, 1 class with 5%, 7 classes with 10% each, and 1 class with the remaining 15% of the observations. The reason for the uneven division was to get a finer distinction between the production classes in the lower part of the spectrum, which was shown to be beneficial in [16]. Before the calculations, the peak yields were adjusted for day in lactation up to day 60, based on a 4th degree polynomial estimated separately for first and later lactations, in the same data. The rationale for this adjustment was that we assumed (and this was corroborated in the data) that before day 60 true peak yield was not achieved even though it was the highest test-day yield reported. The cows were not allowed to have information on peak class before day 10 of lactation, *i.e.* until day 10, all the cows were in the same (dummy) production class and after day 10 they switched to their true class. Cows with missing information on peak yield ($n = 94\,304$ lactations) were set to their class in the previous lactation, or to the dummy class if it was the first lactation

or if the previous lactation peak class was missing. In the Cox analysis this effect was reduced to *par-peak*.

ys-stage-peak

Where year-season, stage of lactation and peak yield classes were defined as described above. In the Cox analysis this effect was reduced to the *ys-peak*.

sire and mgs

Random time-independent effects of transmitting abilities of the cow's sire and maternal grandsire ($0.5 \times$ sire effect, ignored if he was unknown); assumed to follow a multivariate normal distribution with mean zero and variance $A\sigma_s^2$, where σ_s^2 is the variance among sires, and A is the relationship matrix. In the Cox analysis, σ_s^2 was fixed to 0.048.

The three-way interaction between lactation, stage, and peak was included because all cows had the same peak class during the first stage (the first 10 days). This creates a confounding between the dummy peak class and the first stage of lactation. Since the effect of the first stage of lactation might be different in different lactations, *lact-stage-peak* was included. Preliminary analyses showed that the effect of the first stage of lactation was different in different *ys*, and thus the *ys-stage-peak* was included as well.

Heritability was calculated in two ways: (1) $h_1^2 = 4\sigma_s^2/(1 + \sigma_s^2)$ and (2) $h_2^2 = 4\sigma_s^2/(1 + \sigma_{hys}^2 + \sigma_s^2)$ according to [21]. Variance of *hys* (σ_{hys}^2) was the trigamma function evaluated at γ . The trigamma function is the second derivative of the logarithm of the gamma function [14]. To compare models with different numbers of parameters, the Akaike Information Criterion (AIC) was calculated [1] and ([15] p. 386) (comparisons between (non nested) PL and LACTPL models are valid since they are based on exactly the same information). The most adequate model is the one that minimises:

$$\text{AIC} = -2 \ln L + 2p$$

where p is the total number of parameters estimated (*i.e.* number of levels of fixed effects, number of variance components, including the parameters of the baseline hazard function), and where L is the maximum value of the likelihood function (after deleting the sire – maternal grand sire component of the model). All analyses were done using the Survival Kit [9]. Details on how to process stratified frailty models and time-dependent variables with the Survival Kit are given in the User's Manual (<http://www.boku.ac.at/nuwi/software/softskit.htm>).

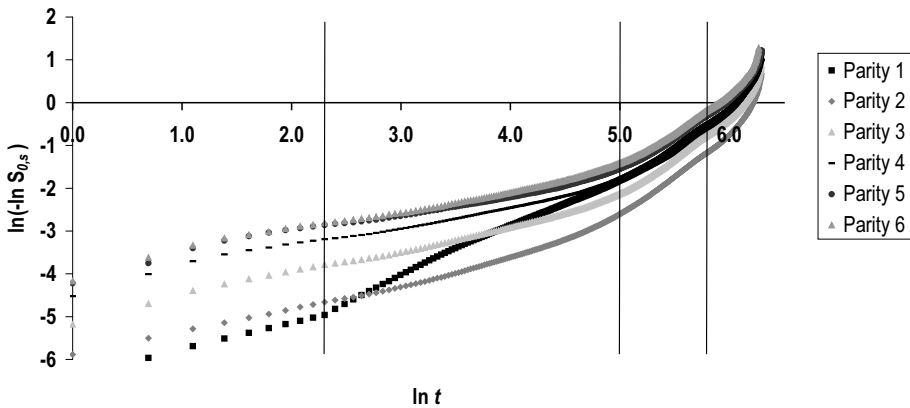


Figure 1. Survivor function ($\ln(-\ln \hat{S}_{0,s}(t))$) versus logarithm of days from previous calving ($\ln t$) from Cox analysis for various strata (lactations). A straight line indicates that a Weibull model can be applied for that lactation; parallel lines indicate that a common baseline across lactations can be used. Vertical lines show where the boundaries for the stage of lactation effect were set in the analyses.

3. RESULTS AND DISCUSSION

3.1. Graphical tests

The plot of $\ln(-\ln \hat{S}_{0,s}(t))$ against $\ln t$ (Fig. 1) suggested that the assumption of a unique Weibull baseline hazard function was not plausible. However, the lines were relatively straight up to $\ln t = 5.0$, but to a lesser extent thereafter. This suggested that the stage of lactation effect probably could not be ignored and that a combination of a Weibull baseline hazard function and a within lactation stage effect would result in a better fit.

Figure 1 shows that the slope of the curve changed somewhere in mid-lactation (around 150 days, $\ln t = 5.0$) and in the end of the lactation (around 300 days). It was not particularly easy to decide, based on the graph, exactly where the stage boundaries should be set. Other boundaries and particularly more stages might be warranted; however, the aim was to use as few stages as possible for LACTPL. Therefore, four different subdivisions of the stage of lactation were tested; either two stages of lactation or three stages of lactation as described previously under the Models (Materials and Methods). When plotting the hazards from the Cox model in each lactation (not shown) it was clearly seen that there was a distinct increase during the first few days after calving, thus justifying including a first stage 0–10 days after calving.

In Figure 1, first lactation differs from later lactations, which are all roughly parallel. This indicated that two different baselines might be enough, one for the first lactation and one for later lactations. Therefore, for LACTPL,

three alternatives were tested: a single baseline hazard function, two baseline hazard functions (first *versus* later lactations) or one baseline hazard function per lactation, *i.e.* one, two, or six strata.

3.2. Model comparisons

The AIC resulting from the analyses of PL and LACTPL with different numbers of strata and between different subdivisions of the stage of lactation effect are summarised in Table I. It was concluded that it was not possible to ignore the stage of lactation effect even if one stratum per lactation, as in LACTPL with six strata, was applied. If only two stages were included, the best definition was 0–150, 151-days of lactation. However, there was added benefit from also including the early stage 0–10 days of lactation. This definition (0–10, 11–150, 151-days of lactation) was actually the best of those studied, regardless of whether it was PL or LACTPL that were analysed and regardless of the number of strata in cases of LACTPL. Some combinations of strata and number of stages which were expected to be very bad given the six strata results were not run in order to save computation time.

There was a clear benefit of having six strata for LACTPL, irrespective of the stage definition used. The reason was probably that the separate baseline for each lactation gave a better fit to the data. However, the benefit of adding strata was less than that of including extra stages in the model.

The sire variance was estimated only for the two models of PL and LACTPL with the lowest AIC. There were minor differences between the models with regards to sire variances, *hys* variances and heritabilities (Tab. II). The estimates of the Weibull parameter ρ in Table II are not directly comparable with each

Table I. Increase in Akaike information criterion (AIC)^a from analysis of longevity based on the entire length of life (PL) and on a lactation basis (LACTPL) compared with the **best model** (with AIC = 5219333).

Stage of lactation	p ^b	PL	LACTPL		
			1 stratum	2 strata	6 strata
0–10, 11–305, 306	847	27 666	11 206	11 631	10 401
0–10, 11–150, 151	861	2011	1403	1305	0
0–150, 151	837	3633	8398	6602	4935
0–10, 11	539	63 530			30 239
No stage effect	543				46 681

^a AIC = $-2 * \ln(\text{maximum value of the likelihood function}) + 2 * \text{total number of estimable parameters (p)}$.

^b For the PL and LACTPL models with one stratum. For the models with 2 and 6 strata, 1 and 5 should be added to p, respectively.

Table II. Parameters from analyses of longevity based on the entire length of life (PL) and on a lactation basis (LACTPL) with stage of lactation effect partitioned at 0–10, 11–150, 151-days of lactation and six strata.

Parameter ^a	Trait	
	PL	LACTPL
ρ^b	1.60 (0.005)	1.45 (0.007); 1.65 (0.010); 1.39 (0.010); 1.25 (0.012); 1.24 (0.017); 1.22 (0.021)
γ	4.83	4.88
σ_{hys}^2	0.23	0.23
σ_s^2	0.043	0.048
h_1^2	0.165	0.183
h_2^2	0.135	0.150

^a ρ : shape parameter of the Weibull distribution (standard error of estimate within parenthesis); γ : parameter of the Gamma distribution; σ_{hys}^2 : herd-year-season variance; σ_s^2 : sire variance; h_1^2 (h_2^2): heritability without (with) herd variance included in the denominator.

^b One per parity (1–6⁺) for LACTPL.

other and with the slopes obtained in Figure 1. As already indicated, to properly compare the hazards at different points in time the value of the baseline hazard function at time t should be combined with the estimates of the time-dependent stage of lactation effect(s) at that time: see [5, 7, 12] for examples. Estimates of heritabilities of approximately 10% are consistent with the literature (reviewed in [19]).

Data structure has a considerable impact when studying the effect of the stage of lactation. Other studies [5, 10] did not find the increased risk of culling in early lactation reported in [2] and in the present study. The likely reason is that in [5, 10], the exact culling date was not available and the status (alive vs. culled) of each animal was only based on test day information: cows actually discarded before their first test-day were considered culled at the end of the previous lactation, and the increased risk at the beginning of lactation could not be found.

The conjecture was that a better modelling of the baseline hazard could be achieved within lactation, and therefore LACTPL models were expected to give a lower AIC than the PL models. However, after having concluded that the exclusion of the stage of lactation effect was unsuitable, we found that the result was dependent on how the stage of lactation effect was partitioned. One reason why LACTPL was not always superior to PL might be that in the Weibull distribution the hazard function invariably starts at zero at time zero. Since the hazard was not negligible early in each lactation in the data, and since

in the LACTPL model there is one hazard function per lactation, the overall fit became worse than with PL where only one hazard function is estimated. It might be worthwhile to consider applying some baseline distribution other than Weibull, *e.g.*, a U-shaped one, or a Cox model (if computationally feasible).

The approach may also lead to simpler data handling, since there will not be any need for reconstruction of the whole life record. The number of elementary records per cow will decrease because some within-lactation effects become time-independent. Analysis on a lactation basis may help clear the way for multiple-trait analysis and time-dependent sire by lactation effects [7].

In conclusion, the analysis of longevity on a lactation basis with a proper accounting for the stage of lactation gave a better fit of the model compared with the conventional analysis of longevity using PL. The reason for this is probably an improved description of the data by the different baseline hazards in different lactations.

REFERENCES

- [1] Akaike H., A new look at the statistical model identification, *IEEE Trans. Automat. Control* 19 (1974) 716–723.
- [2] Beaudreau F., Ducrocq V., Fourichon C., Seegers H., Effect of disease on length of productive life of French Holstein dairy cows assessed by survival analysis, *J. Dairy Sci.* 78 (1995) 103–117.
- [3] Cox D.R., Regression models and life-tables (incl. Disc.), *J. Royal Stat. Soc., Series B*, 34 (1972) 187–220.
- [4] Ducrocq V., An analysis of length of productive life in dairy cattle, Ph.D. Thesis, 1987, Cornell University, NY.
- [5] Ducrocq V., Statistical analysis of length of productive life for dairy cows of the Normande breed, *J. Dairy Sci.* 77 (1994) 855–866.
- [6] Ducrocq V., Survival analysis, a statistical tool for longevity data, in: 48th Annual Meeting of the European Association for Animal Production, 25–28 August 1997, Vienna, Austria.
- [7] Ducrocq V., Topics that may deserve further attention in survival analysis applied to dairy cattle breeding: some suggestions, in: Proceedings of Workshop on genetic improvement of functional traits in cattle, Jouy-en-Josas, 1999, *Interbull Bulletin* No. 21, pp. 181–189, Swedish University of Agricultural Sciences, Dept. of Anim. Breed. and Genet., Uppsala.
- [8] Ducrocq V., Casella G., A Bayesian analysis of mixed survival models, *Genet. Sel. Evol.* 28 (1996) 505–529.
- [9] Ducrocq V., Sölkner J., “The Survival Kit – V3.0”, a package for large analyses of survival data, in: Proceeding of the 6th World Congress on Genetics Applied to Livestock Production, 11–16 January 1998, Vol. 27, University of New England, Armidale, pp. 447–448.
- [10] Ducrocq V., Quaas R.L., Pollack E.J., Casella G., Length of productive life of dairy cows, I. Justification of a Weibull model, *J. Dairy Sci.* 71 (1988) 3061–3070.

- [11] Ducrocq V., Quaas R.L., Pollack E.J., Casella G., Length of productive life of dairy cows, II. Variance component estimation and sire evaluation, *J. Dairy Sci.* 71 (1988) 3071–3079.
- [12] Gröhn Y.T., Ducrocq V., Hertl J.A., Modelling the effect of a disease on culling: an illustration of the use of time-dependent covariates for survival analysis, *J. Dairy Sci.* 80 (1997) 1755–1766.
- [13] Gröhn Y.T., Eicker S.W., Ducrocq V., Hertl J.A., Effect of diseases on the culling of Holstein dairy cows in New York State, *J. Dairy Sci.* 81 (1998) 966–978.
- [14] Kalbfleisch J.D., Prentice R.L., *The statistical analysis of failure time data*, John Wiley and sons, NY, USA, 1980.
- [15] Klein J., Moeschberger M., *Survival analysis: techniques for censored and truncated data*, Springer, NY, USA, 1997.
- [16] Strandberg E., Roxström A., Genetic parameters of functional and fertility determined length of productive life in Swedish dairy cattle, *Anim. Sci.* 70 (2000) 383–389.
- [17] Swedish Dairy Association [Cattle statistics], *Husdjursstatistik 1998/1999 (2000)*, Svensk Mjölk, Box 1146, SE-63180 Eskilstuna.
- [18] Visscher P., Thompson R., Yazdi H., Hill W.G., Brotherstone S., Genetic analysis of longevity data in the UK: present practice and considerations for the future, in: *Proceedings of Workshop on genetic improvement of functional traits in cattle*, Jouy-en-Josas, 1999, *Interbull Bulletin* No. 21, pp. 16–22, Swedish University of Agricultural Sciences, Dept. of Anim. Breed. and Genet., Uppsala.
- [19] Vollema A., Longevity of dairy cows: a review of genetic variances and covariances with conformation, *Anim. Breed. Abstr.* 66 (1998) 781–802.
- [20] Vollema A.R., Van Der Beek S., Harbers A.G.F., De Jong G., Genetic evaluation for longevity of Dutch dairy bulls, *J. Dairy Sci.* 83 (2000) 2629–2639.
- [21] Yazdi M.H., Visscher P.M., Ducrocq V., Thompson R., Heritability reliability of genetic evaluations and response to selection in proportional hazard models, *J. Dairy Sci.* 85 (2002) 1563–1577.

APPENDIX

Let θ be the vector of parameters to be estimated. Assume a proportional hazards model with time-dependent covariates $w(t)$. Consider first the case where PL is analysed. The assumed hazard function at time t is:

$$h(t; w(t)) = h_0(t) \exp\{w'(t)\theta\} \quad (\text{A.1})$$

where $h_0(t)$ is the baseline hazard function. Let L_{PL} be the likelihood function. In the case of random censoring, it has been shown (Kalbfleisch and Prentice, 1980) that:

$$L_{PL} \propto \prod_{i \in \{unc.\}} f(y_i) \prod_{i \in \{cens.\}} S(y_i) \quad (\text{A.2})$$

where y_i is censoring or failure time of animal i . $S(\cdot)$ and $f(\cdot)$ are the survivor function and the density function associated with animal i , $\{\text{unc.}\}$ and $\{\text{cens.}\}$ are the sets of uncensored and censored records. Given the relationships between $h(\cdot)$, $f(\cdot)$, and $S(\cdot)$, (A.2) is more often written as:

$$\ln L_{PL} = \text{constant} + \sum_{i \in \{\text{unc}\}} \ln h(y_i) + \sum_i \ln S(y_i). \tag{A.3}$$

The general formula for $S(\cdot)$ indicates that:

$$S(y_i) = \exp \left\{ - \int_0^{y_i} h(u; w(u)) du \right\}. \tag{A.4}$$

If animal i has a left truncated record with B being the time of left truncation (*i.e.*, the origin point is before the beginning of the time interval considered in the analysis, which occurs at time B for animal i) and y_i still being failure or censoring time, one can show with a derivation similar to the one where (A.2) was obtained, that $S(y_i)$ in (A.3) and (A.4) must be replaced in L_{PL} by:

$$C_i = \exp \left\{ - \int_B^{y_i} h(u; w(u)) du \right\}. \tag{A.5}$$

Now, consider that the PL record of animal i can be partitioned into N_i sub-records (in our case, lactations). Let $b_1 = 0, b_2, \dots, b_{N_i}$ be the beginning of each of these sub-records (each calving date). The last (N_i) sub-record will be considered as left truncated at b_{N_i} and censored or uncensored (as in PL) at $b_{N_i+1} = y_i$. Each previous record will be considered as truncated at b_j (with $b_j = 0$ for the first one) and censored at b_{j+1} .

If all these sub-records are treated separately, still assuming that the proportional hazards model (A.1) is true, then the contribution to $\ln h(y_i)$ in (A.3) will remain unchanged by the partition: it is still the hazard at failure time. The computation of $\ln S(y_i)$ in (A.3) can be decomposed as the sum of elementary contributions, $\ln C_j$, where for all sub-records before the last one:

$$C_j = \exp \left\{ - \int_{b_j}^{b_{j+1}} h(u; w(u)) du \right\}, \tag{A.6}$$

and for the last one:

$$C_{N_i} = \exp \left\{ - \int_{b_{L_i}}^{y_i} h(u; w(u)) du \right\}. \tag{A.7}$$

Note that expression (A.7) does not require the form of $h(u, w(u))$ to be the same for two distinct intervals $]b_i, b_{i+1}[$ and $]b_j, b_{j+1}[$.

Indeed, combining the contributions of all sub-records of animal i (that are conditionally independent, given the covariates $w_i(\cdot)$ of animal i) and noting that:

$$\begin{aligned} \prod_{j=0}^{N_i} C_j &= \prod_{j=0}^{N_i} \exp \left\{ - \int_{b_j}^{b_{j+1}} h(u; w(u)) du \right\} \quad (\text{with } b_{N_i+1} = y_i) \\ &= \exp \left\{ - \sum_{j=0}^{N_i} \int_{b_j}^{b_{j+1}} h(u; w(u)) du \right\} \\ &= \exp \left\{ - \int_{b_0}^{b_{N_i+1}} h(u; w(u)) du \right\}, \end{aligned} \tag{A.8}$$

it is clear that the contribution $\ln S(y_i)$ in (A.3) is also unchanged. The likelihood contribution of each record can be expressed as the product of the contribution all sub-records: these sub-records are independent and therefore are uncorrelated. Finally:

$$\ln(L_{\text{LACTPL}}(\theta)) = \ln(L_{\text{PL}}(\theta)). \tag{A.9}$$

The two approaches for the construction of the likelihood are completely equivalent.