SURVIVAL ANALYSIS FOR LENGTH OF PRODUCTIVE LIFE OF BEEF COWS

Karin Meyer

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

SUMMARY

Survival analysis is applied to records for length of productive life of Angus cows, fitting several proportional hazards models. It is shown that a piece-wise Weibull model with a time-dependent stage effect can model the distinct annual culling pattern reasonably well. Analyses fitting a sire model suggest some genetic variability, with heritability estimates on the logarithmic scale ranging from 6 to 10%. Issues of data quality and the need for whole herd inventory recording are discussed.

INTRODUCTION

Length of productive life (LPL) of cows is an important economic factor in beef cattle production. While genetic evaluation for 'survival' of dairy cows using so-called proportional hazards (PH) models is an integral part of many dairy improvement programmes, few attempts have been made to extend such analyses to beef cattle. PH models imply that each cow which has survived to a given time t has a certain risk to be culled at this time, the hazard $h(t)$. This is modelled as a baseline hazard function, $h_0(t)$, modified by certain risk factors (ω) which can be a combination of fixed and random effects, $h(t) = h_0(t) \exp(\mathbf{w}'\omega)$ (with $\mathbf{w}$ the vector of covariables). PH models are linear on the logarithmic scale, i.e. risk factors act in a multiplicative fashion. A major advantage of PH models is that censored records, i.e records for individuals still alive but likely to be culled in the future, can be included in the analysis. An introduction to survival analyses is given by Kachman (1999). This paper presents a first study applying this methodology to characterise survival of Angus cows in seedstock herds, comparing different models and examining the importance of various risk factors.

MATERIAL AND METHODS

Data. Raw data consisted of pedigree information for 1.66 million Angus cattle, extracted from the National Beef Recording Scheme data base. From these, birth dates of calves were collated for all cows occurring as dams. To accommodate embryo transfer (ET) and cross-fostering, calves were assigned to the rearing rather than the genetic dam. LPL was then defined as the number of months between the birth of the first and last calf on record. Any records less than 10 months were discarded, i.e. any cows without a second calf had LPL= 0 and were not considered. Further edits eliminated any cows with unknown sire or cows born prior to 1980 or after July 2005, and restricted age at birth of the first calf to 20 to 42 months. Records for cows with the last calf born after February 2007 were deemed to be 'censored'. For the analysis, only daughters of sires with at least 5 records were considered, and records for any herds with less than 50 records or less than 40% of cows by a sire used in more than one herd were disregarded. This yielded 98 804 records for daughters of 9 734 sires in 477 herds. Of these, 74.6% were uncensored with a mean LPL of 47.4 months and range of 10 to 174 months. To avoid problems with few records at the higher ages, 638 records above 144 months were set to this value and the records treated as censored. This increased the proportion of censored records to 26.0% while reducing the mean of uncensored observations to 46.5 months.

Analysis. Records were analysed fitting Cox or Weibull PH models, as implemented in the Survival Kit (Ducrocq and Sölkner 1998), by now in Version 5.1. All analyses fitted age of cow at birth of first calf, year of birth of cow, month of birth of cow, month of birth of first calf and month of

birth of last calf as discrete fixed effects. Other risk factors fitted throughout were the number of herds a cow's sire had progeny in $(1, 2 - 5, \geq 6)$, the number of calves which subsequently became sires $(0, 1, \geq 2)$, and whether or not a cow raised an ET calf, was an ET animal herself or changed herds. Herd-year of birth of first calf effects (HY, with 7330 levels) were fitted as either fixed (HY-fix), random (HY-rnd) or random within a fixed herd effect (H+HY-rnd). If treated as random, HY effects were assumed to have a log Gamma

Figure 1. Hazard rates with distribution of records

distribution with equal scale and shape parameter. For Weibull analyses, a time-dependent stage effect with up to 22 classes $(10 - 14, 15 - 22, 23 - 27, 28 - 34, 35 - 39, \ldots, 136 - 144 \text{ months})$ was considered to model changes within and across years. Furthermore, analyses fitting up to 6 time-dependent strata, i.e. a piece-wise Weibull model (Ducrocq 2002), were carried out.

Estimates of genetic variance were obtained fitting a sire model, with sire effects assumed to follow a multivariate normal distribution, accounting for pedigree information on sires and maternal grand-sires (12 681 sires in the pedigree). Pseudo-heritabilities on the logarithmic scale were computed as $h_{\log}^2 = 4\sigma_S^2/(\sigma_S^2 + \sigma_\gamma^2 + \pi^2/6)$ (Ducrocq et al. 1988) with σ_S^2 and σ_γ^2 the variance components due to sires and HY effects, respectively. Values on the original scale were obtained as $h_{\text{org}}^2 = \exp(\nu/\rho)^{-2} h_{\log}^2$, with $\nu \approx 0.5772$ (Euler constant) and ρ the shape parameter of the Weibull distribution, and the 'effective' heritability (Yazdi et al. 2002) was calculated as $h_{\text{eff}}^2 = 4\sigma_S^2/(\sigma_S^2 + \sigma_\gamma^2 + 1)$.

RESULTS

Estimated hazard rates from non-parametric analyses together with the distribution of LPL records are shown in Figure 1. Both exhibit a clear annual pattern, with each year separable into 5 and 7 months periods of high and low risks and numbers of observations, respectively, and highest hazards at approximately 12 monthly intervals. Peak risks of being culled dropped slightly from 12 to 24 months and increased with age after an LPL of 48 months. Estimates of hazard from a Kaplan-Meier analysis (K-M), i.e. not fitting any risk factors, agreed well with those from Cox PH models initially, but were consistently lower 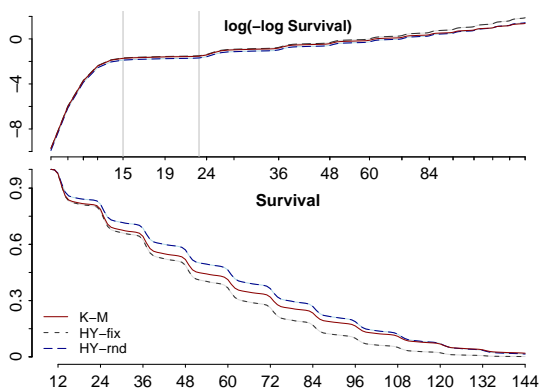at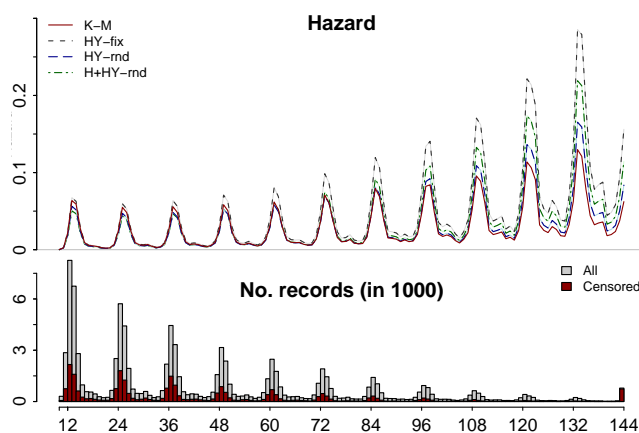 later ages. Fitting HY or herd effects as fixed inflated estimates for the highest ages substantially, suggesting some problems in the data structure due to limited numbers of records. Corresponding survival curves are displayed in Figure 2. Again, annual fluctuations are evident with periods of high and low hazards corresponding to steep and flat parts of the curves. Under the assumption of proportional hazards, a plot of log(− log(Survival)) against log(LPL) is expected to follow a straight line. Clearly this was not the case, with periods from $10 - 14$ and $15 - 22$ months distinctly different and linearity from about 35 months only approximately true.

Figure 2. Non-parametric survival curves
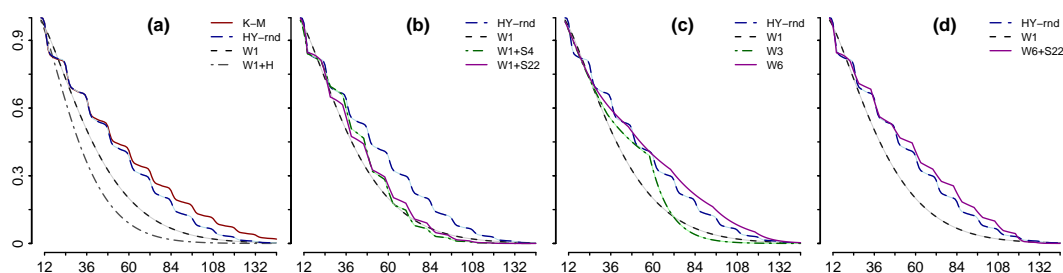
Figure 3. Estimated survival curves from Weibull analyses (see text for definitions).

Estimates of survival curves from different analyses fitting a Weibull hazard function are summarised in Figure 3. All analyses fit HY effects as random. Part (a) contrasts estimates from the non-parametric analyses (*c.f.* Figure 2) with those fitting a single Weibull curve without time-dependent covariables, either with (W1+H) or without (W1) fixed herd effects in the model of analysis. With 46% of records below 28 months, the single curve (W1) agreed reasonably well with the estimate from the Cox model, which does not make any assumptions about the shape of the baseline hazard, in this part but overestimated hazards later on, resulting in a survival curve which declined too quickly. Fitting herds in addition to HY effects exacerbated the deviations. As shown in part (b), allowing for a time-dependent stage effect modelled the wave pattern of the survival curve, but did not increase the quality of fit substantially. Whether this involved separate effects for all 22 stages (W1+S22) determined from the change-points in Figure 2, or whether the model was reduced to 4 stages (W1+S4) by merging odd/even stages from 23 months onwards into 2 effects had rather little impact. Allowing for different Weibull curves in different parts of cows' life, however, dramatically improved the fit of the Weibull model; see part (c). While the log(-log) plot suggested separate curves for records below 15 and between 15 and 23 months, such analyses proved unsuccessful, either due to numerical problems or with the resulting curves 'worse' than the single curve (W1). Fitting three strata (W3) with changes at 35 and 59 months yielded a good fit in the first two parts. However, 6 partial curves (with changes at 23, 47, 71, 95 and 119 months) were required to model the complete curve adequately. Further sub-stratification was attempted, but again failed. As shown in part (d), combining the 6-part Weibull curve with stage effects (W6+S22) yielded a model closely following the estimate from the non-parametric analyses.

Table 1 gives estimates of parameters, variance components and the resulting heritabilities (in %) from Weibull analyses in- and excluding sires, with γ the scale and shape parameter of the log Gamma distribution for HY effects and b_0 the 'intercept' of the curve, $b_0 = \rho \log(\lambda)$ where λ is the scale parameter of the Weibull function. For analyses fitting strata, $\hat{\rho}$, $\hat{b}_0$ and h^2_{org} given pertain to the first stratum. Estimates of the variance due to HY effects were large, ranging from 0.56 to 0.66 for analyses not fitting herds, corresponding to a parameter of the Gamma distribution around 2. Fitting fixed herd effects reduced both $\hat{\sigma}^2_\gamma$ and $\hat{\sigma}^2_S$ in proportion so that heritability estimates were little affected. For all analyses, fitting sires reduced the estimate of $\hat{\sigma}^2_\gamma$ slightly in spite of attempts to reduce confounding of HY and sire effects by restricting data to herds with at least 40% progeny of sires used in more than one herd. Standard errors of $\hat{\sigma}^2_S$ ranged from 0.0038 (W1+H) to 0.0051 (W1+S4). Fitting stages or a piece-wise baseline function increased estimates of σ^2_γ by up to 13% and of σ^2_S by up to 71%, resulting in heritability estimates about 50% higher than those from an analysis fitting a single Weibull curve without stages (W1). In contrast, Ducrocq (2005) found that fitting a more detailed model with a piece-wise curve reduced the estimate of the sire variance for survival of French dairy cows from 0.053 to 0.035. Results from analyses W1 were comparable to values of $\hat{\gamma} = 1.94$ and $\hat{\sigma}^2_S = 0.029$ reported by Forabosco et al. (2006) for Chianina cattle. However,

Table 1. Estimates of parameters from Weibull analyses (see text for definitions)

Model	Without sire effects				Fitting sire effects							
	$\hat{\gamma}$	$\hat{\rho}$	$\hat{b}_0$	$\hat{\sigma}^2_\gamma$	$\hat{\gamma}$	$\hat{\rho}$	$\hat{b}_0$	$\hat{\sigma}^2_\gamma$	$\hat{\sigma}^2_S$	$\hat{h}^2_{\log}$	$\hat{h}^2_{\text{org}}$	$\hat{h}^2_{\text{eff}}$
W1	2.160	1.352	-4.920	0.5861	2.250	1.362	-4.849	0.5569	0.0344	6.2	2.6	8.8
W1+H	4.872	1.367	-4.566	0.2278	5.154	1.375	-4.525	0.2141	0.0283	6.0	2.6	9.1
W1+S4	2.048	2.062	-7.293	0.6263	2.166	2.086	-7.232	0.5842	0.0590	10.3	5.9	14.4
W1+S22	2.058	3.231	-7.352	0.6224	2.164	3.243	-7.248	0.5846	0.0554	9.7	6.8	13.5
W3	2.037	1.098	-4.197	0.6266	2.155	1.104	-4.090	0.5877	0.0531	9.3	3.3	12.9
W6	1.953	1.102	-4.235	0.6645	2.050	1.107	-4.127	0.6252	0.0575	9.9	3.5	13.7
W6+S22	2.063	3.160	-7.226	0.6202	2.167	3.170	-7.116	0.5838	0.0534	9.4	6.5	13.0

Phocas and Ducrocq (2006) obtained estimates of sire variances of 0.04 to 0.05 for LPL of Charolais cattle, but variances due to HY effects of 0.04 to 0.06, differing substantially from our results.

DISCUSSION

Analyses have identified a distinctive, annual pattern for LPL of Australian beef cows with alternating stages of low and high risk, corresponding to periods of rearing a calf and the next calving. In part, this is inherent in our definition of LPL. Results have shown that survival analyses for LPL of beef cows under a Weibull model require careful modelling to adequately represent the baseline hazard function identified in non-parametric analyses, combining a piece-wise Weibull curve with time-dependent stage effects. There is clear evidence of some genetic variation. However, estimates of σ^2_S were small in comparison to environmental variation due to HY effects. Risk factor fitted, other than stage, explained relatively few differences in culling patterns. Disconcertingly, estimates of σ^2_S depended strongly on the model fitted and increased rather than decreased with more detailed modelling. It might be argued that an ill-fitting single curve (analysis W1) obscured differences between sires. Conversely, we rule out that the increase in variances was not, in part at least, due to other, unidentified systematic risk factors or problems inherent in the data structure.

The main limitation in this study has been the lack of 'proper' records for LPL, based on whole herd inventories, in particular recording of disposal dates and reasons. The measure of LPL used is a minimum and relies on registration of all calves to be informative. In addition, it does not account for culling based on failure to raise a calf until weaning. Use of weaning weights records to extract such information was disregarded as the ratio of numbers weaned to born was too low to assume complete, non-selective recording. If a genetic evaluation scheme for survival of beef cows is to be contemplated, it should be accompanied by instigating the appropriate recording scheme.

REFERENCES
Ducrocq, V. (2002) *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.* **19**:20.
Ducrocq, V. (2005) *Anim. Sci.* **80**:249.
Ducrocq, V., Quaas, R. L., Pollak, E. J. and Casella, G. (1988) *J. Dairy Sci.* **71**:3071.
Ducrocq, V. and Sölkner, J. (1998) *Proc. Sixth World Congr. Genet. Appl. Livest. Prod.* **27**:447.
Forabosco, F., Bozzi, R., Filippini, F., Boettcher, P. J., Van Arendonk, J. A. M. and Bijma, P. (2006) *Livest. Prod. Sci.* **101**:191.
Kachman, S. D. (1999) *J. Anim. Sci.* **77**:147.
Phocas, F. and Ducrocq, V. (2006) *CD-ROM Eighth World Congr. Genet. Appl. Livest. Prod.* 03–13.
Yazdi, M. H., Visscher, P. M., Ducrocq, V. and Thompson, R. (2002) *J. Dairy Sci.* **85**:1563.