



**Universitat Autònoma de Barcelona**

**DEPARTAMENT DE CIÈNCIA ANIMAL I DELS ALIMENTS**

**ESTUDI DE LA LONGEVITAT I LA SUPERVIVÈNCIA  
FINS AL DESLLETAMENT EN BOVINS DE MUNTANYA  
DE LA RAÇA *BRUNA DELS PIRINEUS***

**JOAQUIM TARRÉS I FONT**

**Programa de Doctorat en Producció Animal**

**Bellaterra, novembre de 2005**



**Universitat Autònoma de Barcelona**  
**Departament de Ciència Animal i dels Aliments**

El treball d'investigació "**Estudi de la longevitat i la supervivència fins al deslletament en bovins de muntanya de la raça *Bruna dels Pirineus***" ha estat realitzat per Joaquim Tarrés i Font, i es presenta com a requisit per a optar al grau de Doctor.

Bellaterra, novembre de dos mil cinc.

Els directors de Tesi,

El doctorant,

Dr. Jesús Piedrafita Arilla

Dr. Vincent Ducrocq

Joaquim Tarrés Font

---

Aquest treball ha estat realitzat gràcies a una beca del *Programa de Formació de Profesorado Universitario* del *Ministerio de Educación, Cultura y Deporte* en el marc del "*Conveni per al desenvolupament del Programa de Control de Rendiments i de Millora Genètica de la raça Brun dels Pirineus*", finançat pel Departament d'Agricultura, Ramaderia i Pesca de la Generalitat de Catalunya.

## **AGRAÏMENTS**

Als meus directors, Jesús Piedrafita Arilla i Vincent Ducrocq, per haver confiat en mi.

Als meus pares, a la meva germana, als meus oncles i als meus cosins.

Als amics i companys de la Unitat de Genètica i Millora i de la Station de Génétique Quantitative et Appliquée.

A la colla d'Anglès i a la colla de Veterinària.

A les explotacions de La Tria, La Torre, La Borda la Ribera, Casa Vallnova i Mas Fogonella per haver registrat les dades utilitzades en aquest estudi.

Finalment, a totes les persones que d'una manera o altra han contribuït a la realització d'aquest treball.

## **RESUM**

El programa de millora de la raça Bruna dels Pirineus ha de definir i avaluar els caràcters biològics que tenen importància econòmica en la producció. En la primera aproximació a la determinació dels objectius de selecció, els caràcters biològics amb major valor econòmic van ser l'interval entre parts, la prolificitat, la taxa de reforma, la supervivència del vedell del naixement fins a l'escorxador, el pes al deslletament, el creixement durant l'engreix i la conformació de la canal. Dins d'aquests, els caràcters reproductius, particularment la prolificitat, juntament amb la taxa de reforma de les vaques i la supervivència dels vedells en diferents períodes eren les variables amb el major pes econòmic en la funció de beneficis. No obstant, per a expandir i completar el model encara és necessari, entre d'altres, fer un estudi dels diferents factors que afecten la longevitat de les vaques (relacionada amb la taxa de reforma) i la supervivència dels vedells fins al deslletament. L'objectiu d'aquest treball és l'estudi dels factors que afecten la longevitat de les vaques de carn de la raça Bruna dels Pirineus i la supervivència dels seus vedells fins al deslletament mitjançant tècniques d'anàlisi de supervivència. A la vegada, es valida un mètode per estimar correlacions entre caràcters lineals i de supervivència i la seva aplicació a l'avaluació genètica.

El primer capítol del treball presenta una revisió bibliogràfica que dona una visió general de la situació i l'evolució d'aquestes tècniques en la seva aplicació a la producció animal. Les tècniques d'anàlisi de supervivència són les més apropiades per estudiar la longevitat i la supervivència ja que permeten l'ús de tota la informació disponible (completa i censurada) i fan una modelització més correcta dels factors mitjançant models no lineals que incorporen l'ús de variables que modifiquen el seu efecte amb el temps.

En el segon capítol es realitza una anàlisi de la longitud de la vida productiva de les vaques amb una aproximació no paramètrica, la qual dona una mitjana de 9 anys de vida productiva i una taxa de reposició anual de l'11%. Posteriorment, usant un model de riscos proporcionals estratificat per l'explotació, s'estudia la influència del pes al naixement i el guany de pes durant la lactació dels vedells, la dificultat de part, l'interval entre parts i l'edat al primer part sobre el risc de reforma. Dos models són explorats: el primer inclou variables dependents del temps que prenen el valor de la variable en cada data de part; mentre que el segon també comprèn interaccions dependents del temps entre el valor de la variable en el part i la seva mitjana dels últims (fins a) tres parts previs. L'efecte explotació-any és l'efecte més important en els

dos models ( $p < 0,001$ ). Els resultats del primer model mostren que el risc de reforma augmenta amb edats al primer part molt grans ( $p < 0,05$ ), dificultats de part creixents ( $p < 0,001$ ), pesos al naixement dels vedells molt grans ( $p < 0,10$ ), guanys de pes durant la lactació molt petits ( $p < 0,001$ ) i intervals entre parts molt llargs ( $p < 0,001$ ). A més a més, els resultats del segon model mostren que aquests increments també depenen de la seqüència de valors de la mateixa variable en els parts previs. Les vaques amb edat al primer part entre 30 i 33 mesos tenen el mínim risc de reforma, el qual creix a mesura que l'edat al primer part és més tardana ( $p < 0,05$ ). Les vaques amb parts no assistits tenen els valors mínims de risc de reforma. Quan es necessita una forta assistència del ramader en el part, les vaques tenen el doble de risc que en els parts no assistits ( $p < 0,001$ ). Quan es necessita la intervenció del veterinari, amb o sense cesària, el risc de reforma esdevé de sis a vuit vegades superior ( $p < 0,001$ ). Les vaques amb pesos al naixement intermedis tenen el menor risc de ser reformats. Aquest risc augmenta lleugerament quan els pesos al naixement són molt grans ( $p < 0,10$ ). No obstant, aquest augment del risc també depèn de la mitjana dels parts previs. Una vaca que pareix un vedell amb un pes al naixement molt gran té el doble de risc de ser reformat si la mitjana dels pesos al naixement previs és petita o mitjana ( $p < 0,05$ ), però no té efecte si la mitjana prèvia és també molt gran. A més, les vaques amb una mitjana de pesos al naixement molt gran tenen un menor risc de reforma en el següent part si el pes al naixement en aquell part no és molt gran ( $p < 0,05$ ). Pel que fa al guany de pes dels vedells durant la lactació, el risc de reforma de les vaques augmenta lleugerament si els guanys són molt petits ( $p < 0,10$ ). No obstant, els resultats del Model 2 mostren que l'augment del risc també depèn de la mitjana dels parts previs. Així, els guanys de pes molt petits no tenen efecte sobre la reforma si la mitjana de guanys previs és intermèdia o gran però augmenten el risc de reforma de tres a cinc vegades si la mitjana prèvia és petita o molt petita ( $p < 0,05$ ). Finalment, les vaques amb intervals entre parts molt grans tenen un major risc de reforma ( $p < 0,001$ ), especialment si la mitjana dels intervals previs també és molt gran. Aquests riscos de reforma superiors impliquen menors funcions de supervivència, la qual cosa incrementa les taxes de reposició encara que només redueix lleugerament les produccions mitjanes de l'explotació. Per exemple, qualsevol circumstància que faci augmentar la taxa de reposició del 11 al 18% faria disminuir la mitjana del pes al deslletament en 3 kg.

En el tercer capítol, les tècniques d'anàlisi de supervivència s'usen per estudiar la supervivència fins al deslletament dels vedells de carn en les àrees muntanyoses

del Pirineu català. La corba de Kaplan-Meier mostra que la supervivència no és una experiència constant al llarg de tot el període de lactació, ja que la taxa de mortalitat és més pronunciada durant el primer mes de vida. L'anàlisi amb un model de riscos proporcionals mostra que bastants factors influencien en la taxa de mortalitat diària, essent l'efecte de l'exploració-any el més important. A més a més, els vedells nascuts en la primera part del període de parts, de setembre a febrer, tenen el menor risc de mortalitat ( $p < 0,001$ ), mostrant que el risc augmenta a mesura que s'acumulen els naixements. Els vedells nascuts de vaques joves (menys de 1300 dies de vida productiva) tenen un major risc de mortalitat ( $p < 0,05$ ). Els parts no assistits presenten el menor risc de mortalitat i aquest s'incrementa fins a cinc vegades a mesura que els parts es fan més difícils ( $p < 0,001$ ). Aquest risc també tendeix a incrementar-se lleugerament quan els pesos al naixement són petits ( $p < 0,10$ ). En canvi, no es detecta un augment del risc en vedells grans, probablement perquè la dificultat de part també està inclosa en el model. Aquests resultats suggereixen millorar l'ambient en la segona part de període de parts i donar més atenció als vedells nascuts de mares joves. La corba de supervivència s'ajusta amb precisió a una funció paramètrica exponencial a trossos, amb punts de tall als 16 i 32 dies. El menor risc correspon al període de 33-180 dies. El risc en els períodes de 17 a 32 dies i de 1 a 16 dies es veu multiplicat per 7 i 26, respectivament. Els riscos relatius estimats pels diferents factors assumint aquest model exponencial a trossos o un model Weibull dependent del temps eren molt similars als descrits més amunt, confirmant la robustesa del model de Cox, de la mateixa manera que els riscos estimats assumint un model mixt incloent el pare com a efecte aleatori. Les estimacions modals de la variància del pare assumint diferents distribucions base eren properes a 0,3, encara que els errors estàndards eren molt grans (0,4). L'heretabilitat de la supervivència dels vedells al deslletament estimada en l'escala binària arriba a un valor de només 0,037, en part a conseqüència que la supervivència era molt alta (96,9%) en la nostra població. En resum, aquests resultats mostren que l'anàlisi de supervivència és una eina potent per analitzar la corba de mortalitat fins al deslletament dels vedells.

El quart capítol aborda la validació per simulació d'un mètode amb dos passos proposat per Ducrocq i col·l. (2001) per estimar correlacions genètiques i avaluar conjuntament la longevitat i dos caràcters lineals, la qual cosa és computacionalment difícil en grans bases de dades. El problema també és estadístic ja que s'han de combinar caràcters descrits per models lineals amb altres no lineals, en presència de censura, i sense una estructura òbvia de distribució conjunta dels residus. Primer, es

fan anàlisis unicaràcter per obtenir estimacions de la variància genètica i per calcular els pseudo-registres i els seus pesos associats. Aquests pseudo-registres són observacions virtuals lliures de tots els efectes ambientals que poden ser usades en un model BLUP animal per donar els mateixos valors genètics que en l'anàlisi inicial. Combinant aquests pseudo-registres en un model multicaràcter i fixant les variàncies genètiques i residuals als valors estimats en el primer pas és possible obtenir estimacions de les correlacions per AI-REML i valors genètics predits per un BLUP multicaràcter aproximat que utilitza informació directa i indirecta sobre longevitat. Les mitjanes de les correlacions genètiques obtingudes en dades simulades confirmen que els pseudo-registres són útils per estimar correlacions genètiques i residuals tant amb un model pare com amb un model animal. La conveniència d'aquesta aproximació s'estudia simulant diferents heretabilitats per a la longevitat (5% i 20%) i diferents correlacions genètiques entre caràcters (0,20, 0,40, 0,60) i observant que totes les mitjanes de les estimacions de correlacions genètiques eren semblants als valors simulats, fins i tot amb un 90% de censura. Quan aquestes correlacions es tingueren en compte en l'avaluació genètica conjunta, no hi hagué cap guany en la fiabilitat dels valors genètics predits per als toros vells (fins a 30% censurats). No obstant, el guany per als toros joves va ser important (4-5 %) quan la variació genètica era alta, i encara més (fins a un 10 %), quan la variació era baixa. Finalment, hem comprovat la robustesa de l'aproximació en diferents situacions. Quan existeixen correlacions residuals diferents de 0 entre els caràcters, el model pare torna a donar estimacions gairebé no esbiaixades de les correlacions genètiques, mentre que les estimacions del model animal estan sobreestimades. Quan es simula una tendència genètica incorrectament estimada que porta a pseudo-registres esbiaixats, l'anàlisi conjunta incloent un efecte del temps podria corregir adequadament aquest biaix.

## **SUMMARY**

The genetic improvement program of the *Bruna dels Pirineus* beef breed needs to define and evaluate the biological traits that have economic importance on production. In a first approximation to the determination of the selection goals, the biological traits with the highest economic values were calving interval, prolificacy, replacement rate, calf survival from birth to slaughter, weaning weight, weight gain during fattening and carcass conformation. Among these, reproduction traits, especially prolificacy, together with cow replacement rate and calf survival in different periods, were the variables with the highest economic value in the profit function. Nevertheless, to expand and complete the model it was still necessary to study the different factors that affect cow longevity (related to replacement rate) and calf survival up to weaning. The aim of this work is to study the factors affecting longevity of *Bruna dels Pirineus* cows and survival up to weaning of their calves, using survival analysis techniques. Subsequently, a method to estimate genetic and residual correlations between linear and survival traits and its application to genetic evaluation has been validated.

The first chapter presents a literature review that gives the state of the art of the Survival Analysis techniques in their application in Animal Production. Survival analysis techniques are the most appropriate to study longevity and survival because they use all available information (censored and uncensored) and make a more correct modelling of the factors involved throughout nonlinear models, allowing the use of variables that modify their effect over time.

In the second chapter, an analysis of the length of productive life (longevity) was performed through a non-parametric approach, giving an average value of 9 years of productive life, and a corresponding replacement rate of 11%. Using a proportional hazards model stratified by herd, the influence of calf birth weight and weight gain until weaning, calving difficulty, calving interval and age at first calving on length of productive life was studied. Two models were explored: the first one included time-dependent variables taking the current value of the covariate at each calving date, while the second one also comprised time-dependent interactions between the value of the covariate of the current calving and its mean value during the last (up to) three previous calvings. The herd-year effect is the most important in both models ( $p < 0.001$ ). Results from the first model showed that the risk of culling increases with very high ages at first calving ( $p < 0.05$ ), increasing calving difficulties ( $p < 0.001$ ), very large calf birth weights ( $p < 0.10$ ), very small weight gains until weaning ( $p < 0.001$ ), and very long



calving intervals ( $p < 0.001$ ). Furthermore, results from the second model showed that these increases also depend upon the sequence of values for the same covariate in previous calvings. Cows with an age at first calving of between 30 and 33 months had the minimum risk of culling, which increases with a greater age at first calving ( $p < 0.05$ ). Cows with non-assisted calvings had minimum hazard values, which increase as calvings become more difficult. When the farmer's assistance was very strong, cows were at a risk of being culled twice that of cows with non-assisted calvings ( $p < 0.001$ ). When the intervention of the veterinary practitioner was needed, with or without caesaria, the hazard became six to eight times higher ( $p < 0.001$ ). Cows with intermediate calf birth weights had the lowest risk of being culled. This risk slightly increased when calf birth weights were very great ( $p < 0.10$ ). However, the increase in the risk also depended upon the means of previous calvings. A cow having a calf with a very great birth weight had twice the risk of being culled if the previous mean was small or medium ( $p < 0.05$ ), but it had no effect if the previous mean was also very great. Furthermore, when the mean of previous BWs was very large, the risk of being culled decreased if the current birth weight was not very large ( $p < 0.05$ ). In the case of weight gain until weaning, the hazard slightly increased when weight gain until weaning was very small ( $p < 0.10$ ). Nonetheless, results from the second model showed that the increase in hazard also depended on the mean of previous calvings. Very small weight gains had no effect on culling if the previous mean was intermediate or high but it had an important effect if it was small or very small ( $p < 0.05$ ). Finally, cows with very long calving intervals had a higher risk of culling ( $p < 0.001$ ), specially if the mean of previous intervals was also very long. These higher risks of culling implied lower survival functions that increased replacement rates, although it only slightly decreased average performances. For example, any circumstance that may increase the replacement rate from 11% to 18% would decrease the average weaning weight by 3 kg.

In the third chapter, survival analysis techniques were used to analyze survival up to weaning of beef calves in the Pyrenean mountain areas of Catalonia. The Kaplan-Meier curve showed that the survival experience was not constant throughout the lactation period, as the mortality rate was more pronounced during the first month of life. A proportional hazards model analysis showed that several factors influenced the instantaneous mortality rate. Herd-year effect had the strongest influence. Calves born in the first part of the breeding season, from September to February, had the lowest mortality risk ( $p < 0.001$ ), showing that mortality risk increases as births accumulate. Calves from cows younger than 1300 days of productive life had a higher risk of

mortality ( $p < 0.05$ ). Non-assisted calvings presented the smallest risk of mortality, and it increased by up to five times as birth became more difficult ( $p < 0.001$ ). This risk also tended to increase slightly when calf birth weight was small ( $p < 0.10$ ); for bigger calves, no increase of risk was detected, probably because calving difficulty was included in the model. These results suggest improving the environment in the second part of the calving period and paying more attention to births from younger cows. The survival curve fitted a parametric piecewise exponential function very well, with cut points at 16 and 32 days. The lowest risk corresponded to the period of 33-180 days, with the risk for the periods of 17 to 32 days and of 1 to 16 days being multiplied by 7 and 26, respectively. Confirming the robustness of the Cox model, the relative risks estimated for the different factors under this piecewise exponential model or a Weibull time-dependent model were similar to those reported above, as well as to those estimated under a frailty model, including the sire as a random effect. The modal estimates of sire variance under different baseline functions were close to 0.3, although the standard errors were very large (0.4). At weaning, the heritability estimate in the binary scale reached a value of only 0.037 because the survival at weaning was very high (96.9%) in this population. Overall, these results show that survival analysis is a powerful tool to analyse the mortality curve until weaning of beef calves.

The fourth chapter is the validation by simulation of a two step approach proposed by Ducrocq et al. (2001) for the estimation of genetic correlations and the joint evaluation of a nonlinear trait such as longevity and linear traits, which is computationally difficult in large datasets. The problem is also statistical because one has to combine linear and non linear traits, in the presence of censoring, and without an obvious structure of the residual joint distribution. First, univariate analyses are performed to get genetic variance estimates and to compute pseudo-records and their associated weights. These pseudo-records are virtual performances free of all environmental effects that can be used in a BLUP animal model, leading to the same breeding values as in the (possibly nonlinear) initial analyses. Combining these pseudo-records into a multiple trait model and fixing the genetic and residual variances to their values computed during the first step, it is possible to obtain correlation estimates by AI-REML and approximate MT-BLUP predicted breeding values that blend direct and indirect information on longevity. Mean genetic correlations obtained on simulated data confirmed that pseudo-records are useful to estimate genetic and residual correlations both with a sire or an animal model. The suitability of this approach is studied by simulating different heritabilities for longevity (0.05 and 0.20)

and different genetic correlations between traits (0.20, 0.40, 0.60) and observing that all the means of the estimations of genetic correlations were close to simulated values, indeed with a 90% of censoring. When these correlations were taken into account in the joint genetic evaluation, there was no gain in reliability of the estimated breeding values for old bulls (up to 30% censored). However, the gain for young bulls was important (4%-5 %) when the genetic variation was high, and even some more (up to 10%) when the variation was low. Finally, the robustness of the approach in different situations was checked. When nonzero residual correlations exist between traits, a sire model gave nearly unbiased estimates of genetic correlations, while the animal model estimates were biased upwards. When an incorrect genetic trend was simulated to lead to biased pseudo-records, a joint analysis including a time effect could adequately correct for this bias.

## **ÍNDEX**

<b>Introducció</b>		1
<b>Objectius</b>		17
<b>Capítol 1.</b>	<b><i>Anàlisi de supervivència: revisió bibliogràfica.</i></b>	21
<b>Capítol 2.</b>	<b><i>Factors influencing length of productive and replacement rates in the Bruna dels Pirineus beef breed.</i></b>	
<b>Capítol 3.</b>	<b><i>Genetic and environmental factors influencing mortality up to weaning of Bruna dels Pirineus beef calves in mountain areas. A survival analysis.</i></b>	87
<b>Capítol 4.</b>	<b><i>Validation of an approximate approach to compute genetic correlations between longevity and linear traits.</i></b>	111
<b>Discussió</b>		135
<b>Conclusions</b>		159



# ***INTRODUCCIÓ***

## **INTRODUCCIÓ**

### **Importància del boví de carn**

Segons estimacions de Desembre del 2003, el cens total de bestiar boví a la Unió Europea (UE-15) era de l'ordre de 77 milions de caps, amb 31 milions de vaques, de les quals 19 milions eren vaques lleteres i els 12 milions restants corresponien a vaques alletants (Taula 1). França encapçalava el grup de països amb major nombre de vaques alletants, seguida d'Espanya, el Regne Unit i Irlanda. En canvi, la producció de carn de boví a la UE-15, que s'estimava en 7,358 milions de tones al final del 2003 (DARP, 2005), es distribueix de manera diferent, ja que els principals països productors són França, Alemanya, Itàlia, Espanya i el Regne Unit, per aquest ordre. La presència d'Alemanya i Itàlia entre els principals productors és deguda, principalment, a una important producció de vedells provinents de vaques lleteres i a la importació de vedells vius per a engreix, respectivament (Serra, 2001). Per la seva banda, veiem que el sector boví espanyol té un pes relativament important a la UE-15 representant el 8,5% del cens total i el 8,8% de la producció de carn (DARP, 2005). Pel que fa a Catalunya, el sector boví sembla estabilitzat en els 700.000 caps i representa tan sols un 10,6% del cens espanyol i un modest 0,9% del cens total de la UE-15 (DARP, 2005).

El consum de carn de boví a la UE-15 era de 19,8 kg per habitant i any a finals de 2002, mentre que a Espanya el consum mig era lleugerament inferior (16,1 kg per habitant i any) (European Commission, 2004). Durant els darrers 15 anys, el consum de carn de boví a la Unió Europea ha presentat una disminució progressiva. D'entre les diverses causes s'hi troben l'encariment relatiu de la carn de boví respecte d'altres carns (porc i pollastre), els canvis en els costums alimentaris, la preocupació creixent pels temes de salut (diètes lleugeres) i, de manera especial, la temàtica sobre la seguretat alimentària (Aguado i González de Chavarri, 1997). No obstant, el fet més important, sens dubte, fou la crisi de l'encefalopatia espongiforme bovina (EEB o BSE: *bovine spongiform encephalopathy*) l'any 1996, a la que es van unir altres esdeveniments (el més recent dels quals fou el brot de febre aftosa a la primavera del 2001). Per tal de pal·liar la crisi i recuperar la confiança del consumidor, la Comissió Europea ha creat una multitud de mesures destinades a que els consumidors recuperin la confiança en la innocuïtat i, en un sentit més ampli, en la qualitat de la carn de boví. Aquestes mesures consisteixen en establir normes de seguretat més

estrictes i controls més eficaços, fomentar mètodes de producció més sostenibles i permetre la “traçabilitat” total dels animals des del seu origen fins al consum de la seva carn (European Commission, 2004).

**Taula 1.** Cens de bestiar boví a la UE (milers de caps) Desembre de 2003.

País	Total animals	Menys de 12 mesos	Entre 12 i 24 mesos	Més de 24 mesos				
				Mascles	Vedelles	Vaques		
						Total	Lleteres	Carn
Alemanya	13.386	4.099	3.146	107	948	5.086	4.338	748
Àustria	2.052	642	446	25	139	801	558	243
Bèlgica	2.684	715	548	43	309	1.069	572	497
Dinamarca	1.681	549	314	14	106	698	589	109
Finlàndia	978	335	250	8	28	356	328	28
França	19.181	4.975	3.518	471	2.165	8.052	4.012	4.040
Grècia	624	198	101	12	27	286	168	118
Holanda	3.735	1.322	623	23	131	1.636	1.551	85
Irlanda	6.223	1.751	1.480	298	415	2.280	1.136	1.144
Itàlia	6.504	2.008	1.502	79	569	2.346	1.913	433
Luxemburg	185	50	40	4	19	71	41	30
Portugal	1.389	392	218	21	57	702	331	371
Regne Unit	10.504	2.892	2.722	286	690	3.913	2.206	1.707
Suècia	1.553	510	375	23	85	561	404	157
Espanya	6.571	2.227	742	131	307	3.164	1.119	2.045
<b>Catalunya</b>	<b>699</b>	<b>502</b>	<b>31</b>	<b>4</b>	<b>12</b>	<b>149</b>	<b>86</b>	<b>63</b>
<b>EU-15</b>	<b>77.250</b>	<b>22.665</b>	<b>16.025</b>	<b>1.545</b>	<b>5.995</b>	<b>31.021</b>	<b>19.266</b>	<b>11.755</b>

Font. EUROSTAT (03/2003) recollida a Estadística i conjuntura agrària num.194-195 del DARP

### **Sistemes de producció de vaques alletants**

La política agrària comuna (PAC) es va crear amb la intenció de reduir la dependència d'Europa de la importació d'aliments de base, inclosa la carn de boví. Les polítiques de sosteniment de preus de la PAC fomentaven la màxima producció i la seva intensificació. Aquest fet va fer que a mitjans dels anys 80 s'haguessin acumulat excedents enormes (les denominades muntanyes de carn i mantega). El 1992 es produeix una reforma fonamental de la PAC, i l'ajuda a l'agricultura deixa de destinar-



se a augmentar al màxim la producció. Es redueix el sosteniment de preus i es compensen els menors ingressos dels agricultors mitjançant ajudes directes a la renta i s'introdueixen mesures per a estimular l'extensificació i la protecció del medi ambient. Posteriorment, en el paquet de mesures de l'Agenda 2000, la Unió Europea va manifestar la seva voluntat de seguir fomentant els mètodes de producció sostenibles i de vincular-los més estretament a objectius mediambientals. Més recentment, arran dels brots de EEB, es va decidir reduir els incentius a la producció de carn de boví i incrementar les ajudes als productors extensius, perquè cada vegada és més evident que la carn produïda d'aquesta forma és la que busca el consumidor.

Dins d'aquesta política agrària a la Unió Europea, el desenvolupament de l'activitat ramadera a Espanya ha tingut en el passat un cert caràcter dual, imposat per les condicions climàtiques i l'evolució de la demanda. En efecte, el règim de precipitacions (escàs i amb un marcat caràcter estacional) limita notablement la quantitat i la qualitat dels recursos pasturables, constituint un fre natural a l'expansió de les espècies de remugants, fonamentalment el boví, mentre que la pressió de la demanda, conseqüència de l'elevació del nivell de vida, va determinar un notable desenvolupament de la ramaderia intensiva no lligada a la terra, fonamentalment porcí i aus. En els últims anys s'observa un canvi de tendència en aquella orientació dual com a conseqüència d'una major atenció a l'aprofitament dels recursos pasturables i dels problemes sanitaris i mediambientals que, en ocasions, planteja l'explotació ramadera intensiva (MAPA, 2003).

La producció de carn bovina prové de dos sistemes de producció clarament diferenciats. Una part prové de vedells nascuts de vaques lleteres. L'altre prové de vedells nascuts del que es coneix com a vaques alletants, és a dir, vaques mare que donen tota la seva llet al seu vedell ja que l'objectiu primordial és la producció de carn. Dels anys 70 fins ara, el cens del sector boví ha experimentat un notable procés de reestructuració interna, amb la disminució de les vaques d'orientació làctia i la forta expansió dels censos de vaques d'aptitud càrnia (MAPA, 2003; Taula 2). Aquesta expansió s'ha desenvolupat amb major intensitat en les zones de cultius poc productius i amb problemes orogràfics per realitzar labors agrícoles (Buxadé i col·l., 1997). L'explicació d'aquest important increment és doble: per una part s'ha produït una substitució real de les vaques d'aptitud làctia per les d'orientació càrnia, però en molts altres casos, sobretot en races d'aptitud mixta, el canvi ha consistit en abandonar l'orientació lletera i passar a ser alletants per acollir-se al sistema d'ajudes i

destinar la seva producció a la cria de vedells, sense portar llet líquida al mercat ja que deixen de munyir-se.

**Taula 2.** *Evolució del cens del sector boví a Espanya (en milers de caps).*

Any	Total animals	Vaques lleteres	Vaques alletants
1970	4282	1827	566
1987	5094	1793	1015
2001	6411	1182	1895
2003	6571	1119	2045

Font. Hechos y cifras de la agricultura, la pesca y la alimentación en España, publicat pel MAPA (2003).

Els sistemes de producció de vaques alletants estan formats per explotacions ramaderes que inverteixen en la compra o en la cria d'una vaca i esperen obtenir uns ingressos per la venda del seu vedell i d'ella mateixa quan és reformada. Per a que el sistema de producció sigui eficient cal que els animals siguin el màxim de rendibles en el seu medi de producció. És per això que, per adaptar-se a les característiques geoclimàtiques de la zona on es produeix, dins d'aquests sistemes hi ha una multitud de variacions regionals, que inclouen diferents tècniques de ramaderia i diferents races de boví. D'aquesta manera es mantenen sistemes de producció que aprofiten els recursos naturals produint una carn específica de la raça i la zona geogràfica. Aquests diferents sistemes de producció a Espanya es poden agrupar en tres grans grups (Buxadé i col·l., 1997): el sistema de zona humida, el sistema de zones semiàrides i el sistema de muntanya.

El **sistema de zona humida** té el seu origen en les zones de la "Cornisa Cantàbrica" on el creixement continu de l'herba al llarg de l'any facilita l'aparició d'un sistema d'explotació més intensificat. L'abundància de pastura permet utilitzar races amb qualitats excepcionals, altament competitives.

El **sistema de zones semiàrides** té el seu origen en les zones de muntanya i deheses del centre de la Península Ibèrica. L'escassetat de recursos pasturables, amb dos períodes crítics com són l'hivern i el final de l'estiu, fan que la base animal sigui de races autòctones que basen la seva rendibilitat i eficàcia en l'equilibri entre unes bones condicions productives i de qualitat de la carn i la rusticitat per a l'aprofitament de pastures en medis difícils amb l'objectiu de produir un vedell en un interval de poc més d'un any.

El **sistema de muntanya** té el seu origen en les explotacions del Pirineu. Es tracta d'un sistema típic vall - port. A finals de primavera i principis d'estiu els animals pugen a les pastures de muntanya (a més de 1500 msm), on romanen fins a la tardor, quan baixen als prats de les valls. La climatologia durant l'hivern, que redueix dràsticament la producció primària o fa impossible la pastura degut a la neu, condiciona la gestió dels ramats obligant, en alguns casos, a l'estabulació dels animals i a utilitzar aliments emmagatzemats (Villalba i col·l., 2000).

### **La raça de Bruna dels Pirineus**

La *Bruna dels Pirineus* és una raça bovina de carn que s'usa per la producció extensiva en les condicions geoclimàtiques i de maneig de les regions muntanyoses de Catalunya. L'origen de la vaca *Bruna dels Pirineus* es situa a les comarques de la Vall d'Aran, Pallars Sobirà, Pallars Jussà i l'Alta Ribagorça, a la província de Lleida. Prové d'antics encreuaments de la vaca autòctona de les comarques pirinenques catalanes (es postula que era un ecotipus de la Pirenaica) amb bovins procedents de Suïssa de la raça Schwitz o Parda Alpina en espanyol (Mujal, 1998). Aquests animals autòctons van rebre la influència repetida de la Schwitz des de finals del segle XIX. Les primeres fonts escrites que parlen d'importacions de vaques suïsses daten del 1922, però es tenen evidències de que al voltant del 1880 es van fer les primeres importacions, probablement a la Vall d'Aran.

El cens aproximat de la *Bruna dels Pirineus* és de 30.000 vaques de més de 2 anys. Pel que fa a la distribució geogràfica, les comarques amb un cens més important són les comarques pirinenques del Pallars Sobirà, el Pallars Jussà i, en menor quantitat, l'Alta Ribagorça i la Vall d'Aran. Altres comarques amb un cens important són el Ripollès, el Berguedà, Osona i alguns ramats al Solsonès. La Federació de la Vaca *Bruna dels Pirineus* (FEBRUPI) agrupa una part important dels ramaders de la raça. Tot i que fenotípicament existeixen varies línies, es manté la uniformitat genètica per l'existència d'un intercanvi genètic entre les diferents poblacions, amb un flux constant de reproductors (Jordana i Piedrafita, 1993).

Els caràcters generals de la raça, d'acord amb el que estableix el reglament del Llibre Genealògic, són que aquesta està constituïda per animals que pertanyen al tronc Alpí amb proporcions harmòniques i eumètriques, amb tendències longilínies. El seu pelatge és de color bru (marronós - gris fosc) amb diferents tonalitats uniformes i amb degradacions del color al nivell dels ulls, morro, aixella, braguer, ventre, cara

interna de les extremitats i perineu. La pell és rosada o bruna, les mucoses rosades i el musell negre. És una raça rústica, amb bona capacitat de pasturar, de precocitat mitjana, molt ben adaptada a les zones de muntanya, amb gran resistència i facilitat d'adaptació a diferents climes (DARP, 1993).

Originàriament, la *Bruna dels Pirineus* va ser una vaca d'aptitud triple: carn, llet i treball. Més endavant es va convertir en una raça d'aptitud mixta: lletera i càrnia, molt adient per a les zones de muntanya amb pastures estivals. Actualment, la seva explotació és orientada cap a la producció càrnia (Mujal, 1998). El sistema de maneig és extensiu, basat en la utilització directa per part de les vaques dels recursos farratgers, amb una càrrega ramadera de 0,9 UGM/SAU (ha). Depenent de l'època de l'any els animals pasturen diferents zones. A l'estiu s'aprofiten, habitualment de forma comunal, les pastures situades en les zones més altes dels ports de muntanya. A la primavera i a la tardor s'aprofiten les pastures intermitges (sotabosc i cultius abandonats) i les pastures del fons de la vall. La bona adaptació d'aquests animals al medi fa que no hagin de ser estabulats durant una bona part de l'any. De fet, només solen practicar un sistema d'estabulació les ramaderies situades en zones d'alta muntanya, en què l'hivern és més fred, i només durant els dies en què baixen més les temperatures (Mujal, 1998). Els vedells lacten durant 6 mesos aproximadament. El vedell deslletat pot ser un producte final (venut per engreix o per reposició). Es guarden per reposició els animals necessaris per mantenir la grandària del ramat de vaques igualant les taxes de reforma i reposició. Tots els animals no venuts i no guardats per reposició, tant mascles com femelles, passen posteriorment a l'engreix amb pinso comercial fins a un any d'edat.

Les vaques presenten una rusticitat elevada per adaptar-se perfectament al medi en què són criades (zones entre 600 i 1200 m d'altitud) (Mujal, 1998). A l'estiu pugen a ports de muntanya de fins a 2000 m. Les aptituds de les femelles són una edat al primer part de 33 mesos, una fertilitat del 90%, un interval entre parts de 385 dies i una facilitat de part amb 90% de normals, 7% amb forta ajuda del ramader i 3% amb assistència veterinària. La seva bona capacitat maternal es manifesta tant en producció lletera com en la cura del vedell. La docilitat és elevada i molt apreciada pels ramaders.

Els vedells presenten uns índexs de creixement molt bons, comparables als que es poden obtenir fent encreuament industrial amb races càrnies altament especialitzades (Mujal, 1998). Els mascles neixen amb 46,5 kg com a mitjana, essent el pes al deslletament de 243 kg als 185 dies, amb una velocitat de creixement del

naixement al deslletament de 1,06 kg/dia. El creixement mitjà durant l'engreix dels mascles és de 1,63 kg/dia. El pes de la canal és de 330 kg als 381 dies d'edat i la seva composició (avaluada al nivell de la 6<sup>a</sup> costella) és d'un 68,2% de múscul, un 16,3% d'os, un 3,0% de greix subcutani i un 9,7% de greix intermuscular. El gran desenvolupament muscular la situa en la categoria U de l'escala EUROP (Piedrafita i col·l., 2003). L'anàlisi canònic revela que la qualitat de la carn mesurada laboratorialment és intermitja entre les diverses races espanyoles analitzades (Gil i col·l., 2001). La producció d'aquesta carn està emparada per la Denominació de Qualitat Vedella dels Pirineus Catalans (DQ, Reglament Ordre 21-12-1994; DOGC 18-01-1995).

### **Programa de millora genètica a la Bruna dels Pirineus**

Jordana i Piedrafita (1993) van elaborar una primera proposta de programa de millora genètica definint la raça *Bruna dels Pirineus* com una raça càrnia d'aptitud materna. Els objectius generals que es pretenien aconseguir per a les mares eren un grau òptim de fertilitat, unes bones qualitats maternals i bones característiques de rusticitat. Pel que fa al vedell, es va pretendre perseguir dos objectius. Primer, un pes elevat al deslletament, és a dir, la selecció aniria encaminada a escollir els animals que tinguin els millors índexs de creixement. Segon, un pes al naixement mitjà per tal que no hi hagi problemes de dificultat de part. Aquest punt era important perquè els ramats estan sotmesos a condicions extensives, la qual cosa fa que moltes vegades els ramaders no siguin presents al moment del part. Per tal de dur a terme aquest programa, el Departament d'Agricultura, Ramaderia i Pesca de la Generalitat de Catalunya va donar suport al programa de Control de Rendiments i de Millora Genètica de la Raça *Bruna dels Pirineus*, mitjançant un conveni subscrit l'any 1990, el qual s'ha anat renovant anualment des de llavors.

Malgrat que els objectius anteriors són adequats en la primera fase del programa de millora, un programa de millora genètica no ha de pretendre únicament l'increment dels creixements i dels pesos (Alenda, 1994). Una selecció a favor dels pesos o creixements, determina canvis en altres caràcters que poden perjudicar la productivitat dels animals en veure's alterades la fertilitat i la viabilitat del vedell. En alguns casos pot ser que sigui perjudicial per a la productivitat una selecció que incrementi el pes (Alenda, 1994). És per això que tot programa de millora formalment establert hauria de

definir i avaluar els objectius de selecció, és a dir, els caràcters biològics que tenen importància econòmica en la producció.

Els programes de millora en boví de carn tenen dos tipus de caràcters a considerar: caràcters de producció (creixement, morfologia, conversió alimentícia, qualitat de la canal) i caràcters funcionals (fertilitat, facilitat de part, caràcter matern, longevitat). A partir d'aquests caràcters, s'han de definir els objectius de selecció i ponderar-los segons la seva importància econòmica per a introduir-los en els diferents índexs de selecció. Per dur a terme aquesta tasca, en paral·lel al conveni per donar suport al programa de Control de Rendiments i de Millora Genètica, la raça *Bruna dels Pirineus*, entre altres, ha estat objecte d'estudi mitjançant un projecte, finançat pel programa FAIR CT95 0702, que va permetre conèixer l'estructura de les explotacions i dels ramats, la productivitat durant l'engreix i, finalment, fer una primera aproximació a la determinació i valoració dels objectius de selecció que va incloure la major part de la informació disponible en la raça *Bruna dels Pirineus* (Quintanilla i col·l., 2000). El model s'inspirà en els models de Bloch (1996) i Phocas i col·l. (1998) desenvolupats per sistemes de producció de races franceses. Els caràcters biològics amb major valor econòmic (objectius de selecció) van ser: interval entre parts, prolificitat, taxa de reforma, supervivència del vedell del naixement fins a l'escorxador, pes al deslletament, creixement durant l'engreix i conformació de la canal (Quintanilla i col·l., 2000). Dins d'aquests, els caràcters reproductius, particularment la prolificitat, juntament amb la taxa de reforma de les vaques i la supervivència dels vedells en diferents períodes són les variables amb el major pes econòmic en la funció de beneficis (Quintanilla i col·l., 2000).

Tot i que encara manca un estudi més profund, aquest estudi preliminar va permetre detectar algunes mancances d'informació que eren necessàries per expandir i completar el model (Quintanilla i col·l., 2000). Dins d'aquestes possibilitats, en aquest treball ens hem centrat en l'estudi de diferents factors que afecten la longevitat de les vaques (relacionada amb la taxa de reforma) i la supervivència dels vedells fins al deslletament.

### ***Importància econòmica i descripció general de la longevitat***

La longevitat és un caràcter que ha rebut atenció creixent en diferents animals de producció (Yazdi i col·l., 2000). Aquest caràcter està íntimament lligat amb el procés de reforma de les vaques. Històricament, la reforma ha estat dividida en dues categories

principals: voluntària i involuntària. La reforma voluntària és desitjable, ja que aspira a millorar la productivitat total de la ramaderia reformant les vaques poc productives (Rajala-Schultz i Gröhn, 2001). En canvi, la reforma involuntària és generalment percebuda com a indesitjable. Aquesta diferència entre la reforma voluntària i involuntària afecta la interpretació de la longevitat i, en conseqüència, s'ha de distingir entre longevitat real, que es defineix com l'aptitud per evitar la reforma, i longevitat funcional, que és l'aptitud per evitar la reforma involuntària (Ducrocq, 1987). D'acord amb aquesta definició, la longevitat funcional és el caràcter que s'hauria de seleccionar ja que permetria evitar la reforma involuntària incrementant el marge de possibilitats per la reforma voluntària.

El problema de la selecció per longevitat funcional és que no hi ha una mesura directa d'aquest caràcter. Les mesures directes són únicament de longevitat real perquè la reforma voluntària sempre existeix en les poblacions ramaderes en certa magnitud (Essl, 1998). En molts casos, aquesta reforma voluntària d'un animal està fortament relacionada amb la seva producció relativa dins del ramat. Així, una manera d'aproximar la longevitat funcional és analitzar les mesures directes de longevitat real amb una correcció de la producció dins el ramat (Ducrocq i col·l., 1988).

Per tal de mesurar la longevitat real dividirem la longitud de la vida d'una femella en dos períodes de temps: (i) el període costós del naixement fins al primer part i (ii) el període productiu que sol començar el dia del primer part i que finalitza el dia de reforma de l'animal (Essl, 1998). El període que interessa realment allargar és el productiu. En canvi, el període costós no interessa allargar-lo sinó més aviat reduir-lo. Aquest període productiu és la mesura directa de la longevitat i generalment s'anomena "longitud de vida productiva", tot i que es poden usar altres termes diferents com "vida en el ramat" o "nombre de parts" (Essl, 1998). La longitud de la vida productiva es mesura normalment en nombre de dies, mesos o anys, tot i que podria ser mesurada en unitats monetàries gastades o guanyades o en quilograms de llet produïts (Ducrocq, 1997).

Un escurçament de la longitud de la vida productiva d'una vaca suposa un increment de la taxa de reposició. Aquest porta a una major necessitat de vedelles de reposició i, per tant, una major compra o producció d'aquestes. Tot i que en la *Bruna dels Pirineus* predomina la autoreposició, en el cas que es produeixin compres, apart del cost associat a aquestes, el productor ha de incórrer en més despeses relacionades amb l'aclimatació de les noves vedelles i fer front als riscos d'introduir noves malalties. Per tant, un millor coneixement del procés de reforma permetria

obtenir recomanacions pràctiques per incrementar la durada de la vida productiva. L'allargament de la vida productiva d'una vaca suposaria una disminució dels costos de reposició i l'habilitaria per assolir la seva capacitat màxima de producció en arribar a la plena maduresa (Essl, 1998). En aquest context, la inclusió de la longevitat, i/o caràcters relacionats amb ella, com a objectiu de selecció hauria de ser considerat (Quintanilla i col·l., 2000).

### ***Importància econòmica i descripció general de la supervivència dels vedells***

Una explotació de vaques alletants és una empresa en la que el ramader és l'empresari que ha invertit en la compra o en la cria d'una vaca i espera obtenir uns ingressos per la venda d'animals per a vida o per a carn. Cada vaca té unes despeses fixes (instal·lacions, maquinàries, mà d'obra, etc.) i variables (alimentació, medicines, veterinari, etc.), i uns ingressos per la venda dels seus productes que en la ramaderia de carn es redueixen al seu vedell i a ella mateixa quan és reformada (Buxadé i col·l., 1997). La determinació dels ingressos per la venda de vedells no és simple. Aquests depenen de la quantitat de carn produïda anualment, a més del pes del vedell, els caràcters de fertilitat, la viabilitat del vedell (supervivència del vedell) i la qualitat del vedell (quantitat i qualitat de la carn) (Buxadé i col·l., 1997). La fertilitat i la supervivència del vedell fins al deslletament són els caràcters més relacionats amb la quantitat de carn produïda i, per tant, tenen prioritat sobre el pes i la qualitat (conformació i qualitat). Garantida una fertilitat i una viabilitat del vedell, el pes i la qualitat del vedell produeixen un afegit a la productivitat (Buxadé i col·l., 1997).

El concepte de supervivència (viabilitat) és la capacitat d'un animal per romandre productiu al llarg d'un període de temps (Everett, 1976a,b; Smith, 1983). Una major supervivència permet un increment del nombre de vedells venuts sense incrementar significativament els costos d'alimentació de les vaques (Quintanilla i col·l., 2000). Sembla doncs necessari integrar els caràcters de supervivència dels vedells en la definició d'un objectiu de selecció global en boví de carn (Quintanilla i col·l., 2000; Goyache i col·l., 2003). Malgrat la seva importància econòmica pels sistemes de producció de vaques alletants, els estudis genètics i fenotípics sobre supervivència dels vedells són escassos i hi ha poca informació disponible (Cundiff i col·l., 1986; Ray i col·l., 1989; Goyache i col·l., 2003).



### ***Importància de l'avaluació conjunta amb predictors***

Tot i la seva importància econòmica, el problema per integrar els caràcters de longevitat i supervivència en la definició d'un objectiu de selecció global és que mostren heretabilitats bastant baixes, donant lloc a avaluacions genètiques amb precisions baixes per als toros joves i la progenie (Ducrocq i col·l., 2001). Afortunadament, existeixen caràcters més heretables que poden ser usats com a predictors primerencs d'aquests caràcters funcionals. Per exemple, en boví de llet, s'han trobat predictors primerencs de la longevitat funcional en una llarga llista de caràcters morfològics (Visscher i Goddard, 1995, Weigel, 1996, Weigel i col·l., 1998, Druet i col·l., 1999, Larroque i Ducrocq, 2001, Buenger i col·l., 2001). La combinació d'aquestes diferents fonts d'informació en un anàlisi multicaràcter permetria augmentar la precisió de les avaluacions genètiques per longevitat i supervivència.

### **Anàlisi de la longevitat i la supervivència en producció animal**

Per analitzar la longevitat o la supervivència durant un període determinat es disposa d'un conjunt de mesures contínues que descriuen la longitud de l'interval entre un punt origen i un punt final (o esdeveniment d'interès). En el cas de la longevitat de les vaques, el punt origen és la data del primer part i l'esdeveniment és la mort o la reforma de la vaca. En el cas de la supervivència del vedell fins al deslletament, el punt de partida és la data de naixement i el punt final és la mort de l'animal. El problema de treballar amb aquestes mesures contínues és que sovint és inútil o impossible (per exemple, perquè s'han deslletat abans) esperar fins que tots els individus presents en l'estudi hagin estat eliminats de l'explotació per disposar de totes les dades completes i poder començar l'anàlisi. Per tant, al finalitzar l'estudi solen existir individus que no han experimentat l'esdeveniment. Quan això es produeix, tenim informació relativa al temps de vida d'un animal però no el coneixem exactament. Per no haver d'excloure'ls de l'estudi, aquests registres s'utilitzen com a censurats. No obstant, la majoria de mètodes estadístics no permeten treballar amb aquest tipus de registres. Si no disposéssim de mètodes que consideressin aquests registres censurats, llençaríem clarament una gran quantitat d'informació útil. A més, aquesta exclusió podria portar a resultats esbiaixats o models estadísticament ineficients (Allison, 1995). Un mètode per analitzar les dades de supervivència que permet la inclusió de registres d'individus censurats i no censurats és **l'anàlisi de**

**supervivència.** Aquest pot ser descrit toscament com l'art de combinar observacions censurades i no censurades en una anàlisi simple (Kalbfleisch i Prentice, 1980), que en un capítol posterior serà descrit amb detall.

### **Bibliografia**

- Aguado J.A., González de Chavarri E. 1997. Situación actual y perspectivas del subsector vacuno de carne en España. A: *Vacuno de carne: aspectos claves*. C. Buxadé (ed.). Mundi-Prensa, Madrid, pp. 40-54.
- Alenda R. 1994. Objetivo de selección en la raza Avileña-Negra Ibérica. *Bovis* 59,11-26.
- Allison P.D. 1995. *Survival Analysis Using the SAS® System. A Practical Guide*. SAS Institute Inc., Cary, North Carolina.
- Bloch C. 1996. *Determination de l'objectif de selection en la race bovine Limousine*. Diplôme d'Agronomie Approfondie en Sciences et Techniques des Productions Animales. INRA Paris-Grignon, Paris.
- Buenger A., Ducrocq V., Swalve H.H. 2001. Analysis of survival in dairy cows with supplementary data on type scores and housing systems from a region of northwest Germany. *J. Dairy Sci.* 84, 1531-1541.
- Buxadé C. 1997. *Vacuno de carne: aspectos claves*. Mundi-Prensa, Madrid.
- Cundiff L.V., MacNeil M.D., Gregory K.E., Koch R.M. 1986. Between and within-breed genetic analysis of calving traits and survival to weaning in beef cattle. *J. Anim. Sci.* 63, 27-33.
- DARP. 1993. *Races autòctones de Catalunya. Bruna dels Pirineus*. Fulls d'informació. Direcció General de Producció i Indústries Agroalimentàries, DARP, Generalitat de Catalunya. P. 2.
- DARP. 2005. Boví 12-2003. *Estadística i conjuntura agrària*, núm. 194-195.
- Druet T., Sölkner J., Gengler N. 1999. Use of multitrait evaluation procedures to improve reliability of early prediction of survival. *J. Dairy Sci.* 82, 2054-2068.
- Ducrocq V. 1987. *An analysis of length of productive life in dairy cattle*. Ph.D. Thesis, Cornell University, Ithaca, New York, USA.
- Ducrocq V. 1997. Survival analysis, a statistical tool for length of productive life data. A: *Proceedings of the 48<sup>th</sup> EAAP Meeting*, Viena (Austria).
- Ducrocq V., Quaas R.L., Pollak E.J., Casella G. 1988. Length of productive life of dairy cows. 1. Justification of a Weibull model. *J. Dairy Sci.* 71, 3061-3070.

- Ducrocq V., Boichard D., Barbat A., Larroque H. 2001. Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index. A: *Proceedings of the 52nd EAAP Meeting*, Budapest (Hungary), Paper GI.4.
- Essl A. 1998. Length of productive life in dairy cattle breeding: a review. *Livest. Prod. Sci.* 57, 79-89.
- European Commission, 2004. Disponible a: <http://europa.eu.int/comm/agriculture/>.
- Everett R.W., Keown J.F., Clapp E.E. 1976a. Relationship among type, production and stayability in Holstein cattle. *J. Dairy Sci.* 59, 1505-1510.
- Everett R.W., Keown J.F., Clapp E.E. 1976b. Production and stayability trends in dairy cattle. *J. Dairy Sci.* 59, 1532-1539.
- Gil M., Serra X., Gispert M., Oliver M.A., Sañudo C., Panea B., Olleta J.L., Campo M.M., Oliván M., Osoro K., García-Cachán M.D., Cruz-Sagredo R., Izquierdo M., Espejo M., Martín M., Piedrafita J. 2001. The effect of breed-production systems on the myosin heavy chain 1, the biochemical characteristics and the colour variables of Longissimus thoracis from seven Spanish beef cattle breeds. *Meat Sci.* 58, 181-188.
- Goyache F., Gutiérrez J.P., Alvarez I., Fernández I., Royo L.J., Gómez E. 2003. Genetic analysis of calf survival at different preweaning ages in beef cattle. *Livest. Prod. Sci.* 83,13-20.
- Jordana J., Piedrafita J. 1993. Programa de mejora genética de la agrupación racial bovina "Bruna dels Pirineus". A: *V Reunión Nacional de Mejora Genética Animal*. Córdoba: Junta de Andalucía. Consejería de Agricultura y Pesca.
- Kalbfleisch J.D., Prentice R.L. 1980. *The statistical analysis of failure time data*. Wiley, New York.
- Larroque H., Ducrocq V. 2001. Relationships between type and longevity in the Holstein breed. *Genet. Sel. Evol.* 33, 39-59.
- MAPA. 2003. X Censos y producciones ganaderas. *Hechos y cifras de la agricultura, la pesca y la alimentación en España*.
- Mujal M.M. 1998. *Anàlisi demogràfica i genètica de la vaca Bruna dels Pirineus*. Treball de recerca del Programa de Doctorat en Producció Animal. Universitat Autònoma de Barcelona.
- Phocas F., Bloch C., Chapelle P., Bécherel F., Renand G., Ménéssier F. 1998. Developing a breeding objective for a French purebred beef cattle selection programme. *Livest. Prod. Sci.* 57, 49-65.

- Piedrafita J., Quintanilla R., Sañudo C., Olleta J.L., Campo M.M., Panea B., Renand G., Turin F., Jabet S., Osoro K., Oliván M.C., Noval G., García P., García M.D., Cruz-Sagredo R., Oliver M.A., Gispert M., Serra X., Espejo M., García S., López M., Izquierdo M. 2003. Carcass quality of ten beef cattle breeds of the South-west of Europe. *Livest. Prod. Sci.* 82, 1-13.
- Quintanilla R., Milán M.J., Díaz C., Goyache F., Piedrafita J. 2000. Selection objectives in three representative Spanish beef cattle breeds. *FAIR1 CT95 0702 – Final Report*, pp. 55-64.
- Rajala-Schultz P.J., Gröhn Y.T. 2001. Comparison of economically optimized culling recommendations and current culling decisions of Finnish Ayrshire cows. *Prev. Vet. Med.* 49, 29-39.
- Ray D.E., Ituly S.B., Roubicek C.B., Benson C.R. 1989. Pregnancy rate, calf mortality and calving date in unsupplemented Hereford range cows. *Livest. Prod. Sci.* 23, 305-315.
- Serra X. 2001. *La raça bovina Bruna dels Pirineus: Qualitat de la canal i de la carn. Característiques bioquímiques del múscul "Longissimus thoracis" en set races autòctones*. Tesis doctoral, Universitat Autònoma de Barcelona.
- Smith S.P. 1983. *The extension of failure time analysis to problems of animal breeding*. Ph.D. Thesis. Cornell University, Ithaca, NY.
- Villalba D. 2000. *Construcción y utilización de un modelo estocástico para la simulación de estrategias de manejo invernal en rebaños de vacas nodrizas*. Tesis doctoral. Universitat de Lleida.
- Visscher P.M., Goddard M.E. 1995. Genetic parameters for milk yield, survival, workability and type traits for Australian dairy cattle. *J. Dairy Sci.* 78, 205-220.
- Weigel K.A. 1996. Use of correlated trait information to improve the accuracy of early prediction of breeding value for length of productive life. A: Preliminary of the International Workshop on EU Concerted Action Genetic Improvement of Functional Traits in Cattle (GIFT), *Interbull Bulletin* 12, 125-135.
- Weigel K.A., Lawlor T.J., Van Raden P.M., Wiggans G.R. 1998. Use of linear type and production data to supplement early predicted transmitting abilities for productive life. *J. Dairy Sci.* 81, 2040-2044.
- Yazdi M.H., Rydhmer L., Ringmar-Cederberg E., Lundeheim N., Johansson K. 2000. Genetic study of length of productive life in Swedish Landrace sows. *Livest. Prod. Sci.* 63, 255-264.

***OBJECTIUS***

## **OBJECTIUS**

D'acord amb tots el antecedents exposats anteriorment, els objectius plantejats en aquest treball són:

1. Fer una revisió bibliogràfica que dona una visió general de la situació i l'evolució de les tècniques d'anàlisi de supervivència en la seva aplicació a la producció animal.
2. Analitzar amb la metodologia de l'anàlisi de supervivència l'efecte de l'edat de la vaca al primer part, la dificultat de part, el pes al naixement i el guany de pes durant la lactació dels vedells i l'interval entre parts sobre la longitud de vida productiva de les vaques de la raça *Bruna dels Pirineus*, assumint diferents models d'anàlisi de supervivència.
3. Analitzar l'efecte del mes del part, la longitud de la vida productiva de la vaca al part, la dificultat de part i el pes al naixement sobre la taxa de mortalitat diària dels vedells de la raça *Bruna dels Pirineus* i fer una estimació de la variància genètica additiva i l'heretabilitat d'aquest caràcter.
4. Validar per simulació el mètode amb dos passos proposat per Ducrocq i col·l. (2001) per estimar correlacions genètiques i fer una avaluació conjunta de la longevitat (caràcter de supervivència) i dos caràcters lineals en un ampli ventall de situacions (diferents correlacions genètiques, heretabilitats i nivells de censura).



# ***Capítol 1***

***ANÀLISI DE SUPERVIVENCIA: REVISIÓ BIBLIOGRÀFICA***



## **ANÀLISI DE SUPERVIVÈNCIA: REVISIÓ BIBLIOGRÀFICA**

### **Mesures de temps fins a un esdeveniment**

L'anàlisi de supervivència és un conjunt de procediments estadístics que permet modelar i analitzar dades en les quals la variable resposta és un conjunt de mesures positives que descriuen la longitud de l'interval entre un punt origen i un punt final, generalment anomenat "*fallida*". El punt final pot correspondre a l'ocurrència de qualsevol tipus d'esdeveniment com la mort o reforma d'un animal, la recuperació d'una malaltia, l'èxit després d'una inseminació, etc. (Allore i col·l., 2001). Aquesta variable aleatòria és el temps de vida i la designarem com a  $T$ . Quan ens referim a un valor específic que pugui prendre la variable  $T$  el designarem amb la lletra  $t$  (Kleinbaum, 1996). Tot i que existeixen diferents mesures binomials que indiquen si un individu ha superat una determinada edat  $t_0$ , l'ús d'una mesura continua com el temps de vida és preferible per analitzar-la amb models de regressió ja que fa més eficient l'ús de tota la informació disponible no restringint les observacions a un punt definit, simple i arbitrari (Ducrocq, 1997).

### ***Censura i truncament***

Una característica peculiar de l'anàlisi de supervivència és com utilitza la informació incompleta que sorgeix de forma habitual quan s'analitzen variables que mesuren la durada d'un procés. En aquest sentit, es fa distinció entre dades censurades i truncades (veure Klein i Moeschberger, 1997; Ducrocq, 2000).

### **Censura**

El fenomen de la censura es produeix quan tenim informació relativa al temps de vida d'un individu però no el coneixem exactament. Existeixen diferents tipus de dades amb informació parcial que podrien ser utilitzades en una anàlisi. Els principals tipus són censura per la dreta, per l'esquerra i dins d'un interval. Per poder combinar observacions censurades i no censurades s'ha d'assumir la hipòtesi de censura aleatòria (Allison, 1995). Podem imaginar que per cada individu  $i$  hi ha un temps de *fallida*  $t_i$  i un temps de censura  $c_i$ , però només un és observat:  $t_i$  per una observació

no censurada i  $c_i$  per una observació censurada. En el cas de censura aleatòria, s'assumeix que  $t_i$  i  $c_i$  són independents (Ducrocq, 1997). Quan la censura és informativa de futures fallides, les estimacions podrien estar esbiaixades, si les variables mesurades no consideren plenament l'associació entre la censura i futures fallides (Allison, 1995).

Censura per la dreta: És de lluny el tipus més freqüent de censura. La majoria de les dades censurades ho són per la dreta perquè els temps de *fallida* reals han estat tallats pel costat dret de l'interval de temps (Allison, 1995). Aquests temps censurats per la dreta representen el límit inferior del temps de *fallida* (Ducrocq, 1997). Per tant, el temps de *fallida* és igual o superior al temps censurat ( $t_i > c_i$ ). La censura per la dreta es produeix en tres casos (Kleinbaum, 1996):

- a) Fi de l'estudi abans que tots els individus hagin experimentat l'esdeveniment. Per exemple, les vaques que continuen vives al finalitzar l'estudi o els vedells que es deslleten. Aquesta censura s'assumeix aleatòria.
- b) Presència d'individus no observats durant part del període o que s'han perdut durant l'estudi. Per exemple, si fem l'estudi el 2002, les vaques que no tenen data de reforma i que el seu últim part registrat era al 1998. Probablement censura informativa (no aleatòria).
- c) Existència de fallides per altres causes diferents a la d'interès. Per exemple, la mort es produeix per altres causes (per exemple, un accident) i l'animal no pot continuar l'estudi. No es pot donar una bona resposta sobre si la censura és aleatòria o no, depèn de cada cas.

Censura per l'esquerra: També pot succeir que tot el que sabem sobre el temps de *fallida* és que és inferior o igual a algun temps censurat ( $t_i < c_i$ ). Per exemple, si un vedell que ha nascut l'1 d'Abril i al voler deslletar-lo l'1 d'Octubre no hi és, sabem que ha viscut menys de 183 dies. Ara parlem de censura per l'esquerra que representa el límit superior del temps de *fallida* real (Ducrocq, 1997).

Censura dins un interval: Un temps de *fallida* és censurat dins un interval ( $t_i \in (c_d, c_e)$ ) si només sabem que l'individu ha experimentat l'esdeveniment entre dos temps de censura: un per la dreta  $c_d$  i un per l'esquerra  $c_e$  (Cantor, 1999). Per exemple, un

vedell que sabem segur que va ser pujat a les pastures de muntanya  $c_d$  dies després del seu naixement i que en baixar les vaques  $c_e$  dies després no hi és.

### **Truncament**

El truncament és una condició que filtra la presència de certs individus de forma que l'investigador no en sap la seva existència. És a dir, els individus participen en l'anàlisi si es produeix la condició de truncament (per exemple, l'exposició a una malaltia, l'ocurrència d'un esdeveniment intermedi, etc.). Tota la inferència s'haurà de fer condicionada a aquesta condició. Existeixen dos tipus de truncament:

**Truncament per l'esquerra:** Es dona quan només s'inclouen en l'estudi els individus que compleixen la condició de truncament amb anterioritat a la *fallida*. Per exemple, les vaques s'inclouen a l'estudi si han tingut un part (qualsevol part) després d'una data determinada.

**Truncament per la dreta:** Es dona quan només s'inclouen en l'estudi els individus que experimenten la *fallida*. Per exemple, una anàlisi de supervivència basat en els vedells amb data de baixa registrada i sense tenir en compte els vedells censurats.

### **Distribució de probabilitat**

El temps de vida és una variable aleatòria que segueix una determinada distribució de probabilitat. Hi ha molts models diferents per tractar les dades d'aquesta variable aleatòria i el que sovint distingeix un model de l'altre és el coneixement de la seva distribució de probabilitat. Existeixen diferents funcions per descriure una distribució de probabilitat continua de manera equivalent (Allison, 1995). Les funcions de distribució usuals per descriure una distribució de probabilitat continua en estadística són la funció de distribució  $F(t)$  i la funció de densitat de probabilitat  $f(t)$ . Suposant que el temps de vida és una variable aleatòria que designem com a  $T$  i que quan ens referim a un valor específic que pugui prendre la variable  $T$  el designem amb la lletra  $t$ , la **funció de distribució  $F(t)$**  en el moment  $t$  es defineix com la probabilitat que un individu sobrevisqui menys de  $t$  anys:

$$F(t) = P(T < t) \quad [1]$$

Teòricament la funció de distribució, definida per  $t \geq 0$ , segueix una corba creixent que comença a  $F(t) = 0$  quan  $t = 0$ , i creix cap a  $F(t) = 1$  quan  $t \rightarrow \infty$  (Kleinbaum, 1996).

La **funció de densitat de probabilitat  $f(t)$**  és la probabilitat no condicionada que un individu experimenti l'esdeveniment en l'instant  $t$ :

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad [2]$$

Es tracta d'una funció no negativa amb integral total igual a 1.

No obstant, existeixen funcions específiques de l'anàlisi de supervivència relacionades amb aquestes funcions de distribució usuals que són generalment més convenients (Ducrocq, 1997). Aquestes funcions específiques són la funció de supervivència  $S(t)$ , la funció de risc  $h(t)$  i la funció de risc acumulat  $H(t)$ . La **funció de supervivència  $S(t)$**  en el moment  $t$  es defineix com la probabilitat que un individu sobrevisqui més de  $t$  anys:

$$S(t) = P(T \geq t) = 1 - F(t) \quad [3]$$

Teòricament la funció de supervivència, definida per  $t \geq 0$ , segueix una corba decreixent que comença a  $S(t) = 1$  quan  $t = 0$  i decreix cap a  $S(t) = 0$  quan  $t \rightarrow \infty$ . La **funció de risc  $h(t)$**  és la taxa instantània d'esdeveniment en el moment  $t$ , condicionada a que l'individu ha sobreviscut més de  $t$  dies,  $T \geq t$ :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \quad [4]$$

El risc no és una probabilitat sinó la taxa d'esdeveniments per unitat de temps i pot prendre qualsevol valor entre 0 i  $\infty$  (Kleinbaum, 1996). En producció animal, el risc és, per exemple, la taxa instantània de reforma a una determinada vida productiva  $t$  (Ducrocq i col·l., 1988a). Si el temps és mesura en dies seria la taxa de reforma diària. Finalment, la **funció de risc acumulada  $H(t)$**  és la suma del risc previ al temps  $t$ :

$$H(t) = \int_0^t h(u) du \quad [5]$$

Tot i que és molt útil gràficament i també tècnicament, no té interpretació intuïtiva.

Totes aquestes funcions usuals i específiques estan interrelacionades. En particular, la relació entre les funcions de supervivència, de risc i de densitat és:

$$S(t) = \exp\left[-\int_0^t h(u) du\right] = \exp[-H(t)] \quad [6]$$

$$h(t) = -\frac{d\log S(t)}{dt} \quad [7]$$

$$f(t) = -\frac{dS(t)}{dt} = h(t) S(t) \quad [8]$$

Per tant, aquestes funcions descriuen la distribució de probabilitat de manera equivalent ja que un pot anar fàcilment d'una a altra.

### **Estimació de la distribució del temps de vida**

Un cop definides les possibles funcions de distribució, un dels objectius de l'anàlisi de supervivència és estimar i especificar els models de distribució del temps de vida. Existeixen diferents mètodes per assolir aquest objectiu que s'agrupen en mètodes empírics (o no paramètrics) i mètodes paramètrics. Tots aquests mètodes assumeixen la hipòtesi de censura aleatòria ja que, si la censura fos informativa de futures fallides, les estimacions podrien estar esbiaixades (Allison, 1995). A partir d'ara només es considerarà la censura per la dreta. La informació sobre com es poden tenir en compte els altres tipus de censura i el truncament es pot trobar a Klein i Moeschberger (1997).

### ***Estimació empírica (o no paramètrica)***

Els mètodes empírics permeten estimar la funció de supervivència del temps de vida sense assumir una distribució paramètrica coneguda. El mètode empíric més emprat és el mètode de Kaplan-Meier (Kaplan i Meier, 1958), tot i que n'existeixen d'altres, com per exemple el mètode actuarial. Suposem que  $t_{(1)} < t_{(2)} < \dots < t_{(k)}$  són  $k$  temps de vida no censurats diferents, ordenats (del més curt al més llarg). Un estimador empíric de la funció de supervivència  $S(t)$  pot ser obtingut descrivint la definició probabilística de  $S(t)$  com un producte de probabilitats condicionals:

$$\hat{S}(t_{(i)}) = P(T \geq t_{(i)}) = P(T \geq t_{(1)})P(T \geq t_{(2)} | T \geq t_{(1)}) \dots P(T \geq t_{(i)} | T \geq t_{(i-1)})$$

En cada moment  $t_{(j)}$  existeixen  $n_j$  individus a risc, que no han estat reformats ni han estat censurats abans de  $t_{(j)}$ . Si  $d_j$  és el nombre d'individus que són reformats en el moment  $t_{(j)}$ , un estimador natural de cada probabilitat condicional és:

$$P(T \geq t_{(j)} | T \geq t_{(j-1)}) = \frac{n_j - d_j}{n_j}$$

Per tant, l'estimador de Kaplan-Meier de la funció de supervivència empírica en cada temps de vida  $t_{(i)}$  es defineix com:

$$\hat{S}(t_{(i)}) = \hat{S}_{KM}(t_{(i)}) = \prod_{j=1}^i \left[ \frac{n_j - d_j}{n_j} \right] \quad [9]$$

Per la seva banda, un estimador no paramètric de la funció de risc en cada temps de vida  $t_{(i)}$  es defineix com:

$$\hat{h}(t_{(i)}) = \frac{d_i}{n_i} \quad [10]$$

O sigui, la proporció del nombre d'individus a risc que són reformats en el moment  $t_{(j)}$ .

### **Estimació paramètrica**

Els mètodes paramètrics permeten estimar per màxima versemblança els paràmetres d'una distribució coneguda per al temps de vida. L'estimació per màxima versemblança comença suposant que per cada individu  $i$  hi ha un temps de vida  $t_i$  i un temps de censura  $c_i$ , però el que realment observem és la variable  $y_i = \min(t_i, c_i)$  i el seu indicador de censura  $\delta_i = \text{Ind}(t_i < c_i)$ . Els temps de vida  $t_i$  tenen una distribució paramètrica coneguda amb funció de densitat  $f(t_i)$  i funció de supervivència  $S(t_i)$ . Els temps de censura  $c_i$  tenen una distribució paramètrica que pot ser desconeguda amb funció de densitat  $g(c_i)$  i funció de supervivència  $G(c_i)$ .

### **Construcció de la funció de versemblança**

Per construir la funció de versemblança presentem el plantejament de Ducrocq (2000). Si l'individu  $i$  té registrada una dada completa ( $\delta_i = 1$ ) al temps  $y_i$  (es compleix que  $t_i = y_i$  i que  $t_i < c_i$ ), la seva contribució a la versemblança és:

$$L_i = \lim_{\Delta t \rightarrow 0} \frac{P(y_i \leq T < y_i + \Delta t)}{\Delta t} P(C \geq y_i) = f(y_i) G(y_i)$$

En canvi, si l'individu  $i$  té registrada una dada censurada ( $\delta_i = 0$ ) al temps  $y_i$  (es compleix que  $c_i = y_i$  i que  $t_i > c_i$ ), la seva contribució a la versemblança és:

$$L_i = \lim_{\Delta t \rightarrow 0} \frac{P(y_i \leq C < y_i + \Delta t)}{\Delta t} P(T \geq y_i) = g(y_i) S(y_i)$$

En ambdós casos, s'ha assumit que les distribucions  $t_i$  i  $c_i$  són independents (censura aleatòria). Per tant, si denotem com  $\varphi$  el conjunt de paràmetres a estimar de la distribució coneguda del temps de vida  $t_i$ , la funció de versemblança de  $\varphi$  és:

$$L(\varphi) = \prod_{i \in \{comp\}} [f(y_i) G(y_i)] \prod_{i \in \{cens\}} [g(y_i) S(y_i)]$$

on  $\{comp\}$  i  $\{cens\}$  representen els conjunts d'observacions completes i censurades respectivament. No obstant, degut a l'assumpció de censura aleatòria, la funció de densitat  $g(\cdot)$  i de supervivència  $G(\cdot)$  dels temps de censura  $c_i$  no depenen dels paràmetres  $\varphi$ . Llavors:

$$L(\varphi) \propto \prod_{i \in \{comp\}} [f(y_i)] \prod_{i \in \{cens\}} [S(y_i)] = \prod_i [f(y_i)]^{\delta_i} [S(y_i)]^{1-\delta_i}$$

i com que  $f(y_i) = h(y_i) S(y_i)$  :

$$L(\varphi) \propto \prod_i [h(y_i)]^{\delta_i} S(y_i) \quad [11]$$

o, prenent logaritmes:

$$\log L(\varphi) = ct + \sum_i [\delta_i \log h(y_i)] + \sum_i \log S(y_i) \quad [12]$$

Els estimadors de màxima versemblança de  $\hat{\varphi}$  s'obtenen maximitzant aquesta expressió.

### **La distribució de Weibull**

Les distribucions clàssiques més usades en l'anàlisi de supervivència són la distribució exponencial, la distribució de Weibull, les distribucions Gamma i Gamma generalitzada, la distribució Log-normal i la distribució Log-logística (Ducrocq, 2000). Klein i Moeschberger (1997) presenten una taula amb les funcions de risc, supervivència, densitat i esperança associades a les distribucions paramètriques més

utilitzades en l'anàlisi de supervivència, que inclou les mencionades anteriorment i algunes més, com la distribució de Gompertz, Pareto, Gaussiana inversa, Exponencial potencial i F generalitzada. Dins de totes aquestes distribucions, destaquen la distribució exponencial amb la seva funció de risc constant  $h(t) = \lambda$  i la distribució de Weibull, que és una generalització amb dos paràmetres  $(\lambda, \rho)$  de la distribució exponencial, amb funció de risc  $h(t) = \lambda\rho(\lambda t)^{\rho-1}$ . La distribució de Weibull ha estat el model paramètric més popular en la literatura bioestadística (Allison, 1995). Una de les raons és perquè té una funció de supervivència relativament simple que és fàcilment tractable matemàticament,  $S(t) = \exp\left[-(\lambda t)^\rho\right]$ . Malgrat que només té dos paràmetres  $\lambda$  i  $\rho$ , el paràmetre  $\rho$  introdueix diferències qualitatives en la forma de la funció de risc que li permet adaptar-se a diferents situacions de dades. La seva interpretació és que quan  $\rho = 1$ , el risc és constant al llarg del temps i les mesures segueixen una distribució exponencial; quan  $\rho < 1$ , el risc decreix amb el temps; quan  $1 < \rho < 2$ , el risc creix a un ritme decreixent; quan  $\rho > 2$ , el risc creix a un ritme creixent; i, finalment, quan  $\rho = 2$ , la funció de risc és una recta creixent amb l'origen a 0 (Allison, 1995).

A l'hora de construir la funció de versemblança assumint una distribució de Weibull definim  $\varphi = (\lambda, \rho)$ . Si tenim  $N = \sum_i \delta_i$  observacions no censurades, la funció de log-versemblança és:

$$\begin{aligned} \log L(\lambda, \rho) &= ct + \sum_i \left[ \delta_i \log(\lambda \rho (\lambda y_i)^{\rho-1}) \right] + \sum_i \log(\exp[-(\lambda y_i)^\rho]) = & [13] \\ &= ct + N \log \rho + N \rho \log \lambda + (\rho - 1) \sum_i [\delta_i \log y_i] - \lambda^\rho \sum_i y_i^\rho \end{aligned}$$

Els estimadors de màxima versemblança de  $\lambda$  i  $\rho$  s'obtenen maximitzant aquesta expressió.

### **Validació de les distribucions paramètriques**

L'estimació paramètrica permet precisar i simplificar les estimacions dels paràmetres però hem d'assegurar que assumim la distribució correcta (Ducrocq i col·l., 1988a). Una aproximació clàssica per escollir una forma adequada per la funció de distribució d'una determinada variable és estimar la seva funció de distribució empírica i comparar



la seva forma amb la de distribucions paramètriques conegudes (Ducrocq, 1997). L'estimador de la funció de supervivència de Kaplan-Meier  $\hat{S}_{KM}(t_{(i)})$  juga un paper clau quan es vol contrastar si la distribució de la variable temps de vida pertany a una família paramètrica particular (Ducrocq, 1997).

### **Validació de la distribució de Weibull: el mètode dels logaritmes**

En el cas especial de la distribució de Weibull, la funció de supervivència paramètrica és:

$$S(t) = \exp\left(-(\lambda t)^\rho\right)$$

i, per tant, prenent logaritmes:

$$\log(-\log S(t)) = \rho \log \lambda + \rho \log t \quad [14]$$

En conseqüència, si la representació del logaritme del logaritme negatiu de la funció de supervivència de Kaplan-Meier  $\log(-\log \hat{S}_{KM}(t))$  en funció del logaritme del temps de vida  $\log t$  descriu una línia recta, es pot concloure que una distribució de Weibull pot ser especificada per ajustar les dades (Ducrocq, 1997). El pendent d'aquesta recta és el paràmetre  $\rho$  i el terme independent és  $\rho \log \lambda$ .

### **Models de regressió**

Fins ara ens hem centrat en l'anàlisi de la distribució del temps de vida en una única població, considerant que no podem diferenciar els individus en funció d'unes variables explicatives. No obstant, a vegades aquesta distribució pot dependre d'un conjunt de variables explicatives. Com a conseqüència, els mètodes presentats fins ara són insuficients per valorar la relació d'una combinació d'efectes diferents amb el temps de vida i s'han d'utilitzar models de regressió (Allison, 1995).

#### ***Model Lineal Normal***

Aquest problema de relacionar un conjunt de variables explicatives amb una variable resposta és anàleg al problema de la regressió lineal. Si  $t_i$  és una variable aleatòria

que denota el temps de vida de l'individu  $i$  en una mostra i que depèn d'un conjunt de variables explicatives  $\mathbf{x}_i'$ , el model de regressió lineal normal seria:

$$t_i = \mu + \mathbf{x}_i' \boldsymbol{\beta} + \sigma \varepsilon_i \quad [15]$$

on  $\boldsymbol{\beta}$  seria el vector de coeficients de regressió a estimar. Els errors aleatoris  $\varepsilon_i$  seguirien una distribució normal amb una esperança i variància constants al llarg de  $i$ , essent independents entre observacions. El paràmetre  $\sigma$  permet fixar la variància de  $\varepsilon_i$  a algun valor estàndard (per exemple, 1) i deixar que canviï el valor  $\sigma$  per acomodar els canvis en la variància de l'error (Allison, 1995). L'efecte de les variables explicatives,  $\mathbf{x}_i' \boldsymbol{\beta}$ , actuaria additivament sobre l'esperança del temps de vida  $\mu$ .

El problema d'aquest model de regressió és que no permet analitzar una variable resposta amb valors restringits a mesures sempre positives com és el cas del temps de vida  $t \geq 0$ . A més, una distribució simètrica com la normal no aproximaria correctament la distribució del temps de vida ja que aquesta sol ser extremadament asimètrica (Vukasinovic, 1999). Per tant, el model lineal normal no seria vàlid per analitzar el temps de vida i haurem d'utilitzar els models de regressió específics de l'anàlisi de supervivència.

### **Model de Temps de Fallida Accelerat**

Un model de regressió específic de l'anàlisi de supervivència és el Model de Temps de Fallida Accelerat (**AFTM: Accelerated Failure Time Model**). Aquest model és el més intuïtiu ja que és similar al model lineal normal. Si  $t_i$  és una variable aleatòria que denota el temps de vida de l'individu  $i$  en una mostra i que depèn d'un conjunt de variables explicatives  $\mathbf{x}_i'$ , el model AFTM (Kalbfleisch i Prentice, 1980) és:

$$t_i = \exp(\mu + \mathbf{x}_i' \boldsymbol{\beta} + \sigma \varepsilon_i) \quad [16]$$

on  $\mathbf{x}_i' \boldsymbol{\beta}$  seria el vector de coeficients de regressió a estimar i l'error aleatori  $\varepsilon_i$  seguiria una distribució determinada. Novament, el paràmetre  $\sigma$  permet fixar la variància de  $\varepsilon_i$  a algun valor estàndard i deixar que canviï el valor  $\sigma$  per acomodar els canvis en la variància de l'error (Allison, 1995). Transformant logarítmicament l'equació, el model s'expressa d'una manera alternativa:

$$\log t_i = \mu + \mathbf{x}_i' \boldsymbol{\beta} + \sigma \varepsilon_i \quad [17]$$

Amb aquesta transformació s'observa que els models AFTM poden ser també descrits com una classe particular de models de regressió log-lineals (Ducrocq, 1997).

Les diferències entre el model AFTM i el model de regressió lineal normal són que: (i) l'efecte de les variables explicatives està exponenciat,  $\exp(\mathbf{x}_i' \boldsymbol{\beta})$ , per assegurar que els valors predits de  $t_i$  siguin sempre positius (l'efecte de les variables explicatives actua proporcionalment sobre el temps de vida representant una acceleració en l'escala d'aquest temps quan  $\exp(\mathbf{x}_i' \boldsymbol{\beta}) > 1$  o una desacceleració quan  $\exp(\mathbf{x}_i' \boldsymbol{\beta}) < 1$ ) i (ii) el model AFTM permet assumir altres distribucions per  $\varepsilon_i$ , apart de la normal, mantenint els supòsits de esperança i variància constants i d'independència entre observacions (Allison, 1995). Novament, les distribucions clàssiques més usades per  $\varepsilon_i$  en l'anàlisi de supervivència fan que la distribució del temps  $t_i$  sigui l'Exponencial, la distribució de Weibull, la distribució Gamma o la distribució Log-logística.

### **Model de Riscos Proporcional**

Tot i que el model AFTM és més intuïtiu, el model de regressió específic de l'anàlisi de supervivència més popular és el Model de Riscos Proporcional (**PHM: Proportional Hazard Model**). Aquest model utilitza la funció de risc, que és una funció específica de l'anàlisi de supervivència per descriure la probabilitat de distribució del temps de vida. Així, el model PHM (Cox, 1972) assumeix que la funció de risc d'un individu  $h(t_i; \mathbf{x}_i)$  s'explica per un producte de dos termes:

$$h(t_i; \mathbf{x}_i) = h_0(t_i) \exp(\mathbf{x}_i' \boldsymbol{\beta}) \quad [18]$$

El primer terme  $h_0(t_i)$  és l'anomenada funció de risc base del temps de vida, que representa el procés d'envelliment de la població (Ducrocq, 1997). El segon terme depèn d'un conjunt de variables explicatives  $\mathbf{x}_i'$ , on  $\boldsymbol{\beta}$  és el vector de coeficients de regressió, que són els paràmetres a estimar. Aquest segon terme està exponenciat,  $\exp(\mathbf{x}_i' \boldsymbol{\beta})$ , per assegurar que els riscos estimats siguin sempre positius. Per tant, l'efecte de les variables explicatives actuarà proporcionalment sobre la funció de risc base  $h_0(t_i)$  que, vagament parlant, seria com una mitjana (Ducrocq, 1997).

### La funció de risc base

Els models de riscos proporcionals permeten assumir una funció de risc base coneguda o considerar-la totalment arbitrària. D'aquesta manera podem classificar-los en models paramètrics o semiparamètrics.

### **Models paramètrics: el model de Weibull**

Per poder estimar els coeficients  $\beta$  de les variables explicatives en models paramètrics, la funció de risc base  $h_0(t_i)$  ha d'estar completament especificada, excepte pels valors dels paràmetres desconeguts de cada distribució. Una vegada especificada la distribució és relativament senzill deduir, calcular i maximitzar la funció de versemblança completa usant les contribucions dels registres complets i censurats (Ducrocq, 1997). Això proveeix estimacions per màxima versemblança dels paràmetres de  $h_0(t_i)$  i dels coeficients de  $\beta$ . El model paramètric més popular és el model de Weibull:

$$h(t_i; \mathbf{x}_i) = \lambda \rho (\lambda t_i)^{\rho-1} \exp(\mathbf{x}_i' \beta) \quad [19]$$

Aquest model de regressió, que és un cas particular de model de riscos proporcionals, és l'únic que també pertany a la família dels models de temps de *fallida* accelerat. Existeix una relació exacta ( $\beta_{PHM} = -\rho \beta_{AFTM}$ ) entre els coeficients estimats pel model de riscos proporcionals (PHM) i pel model de temps de *fallida* accelerat (AFTM), la qual permet mesurar l'efecte proporcional de les variables sobre el temps de vida que és més fàcil d'interpretar que sobre el risc (Allison, 1995).

Seguint Ducrocq (2000), la funció de risc del model de Weibull es pot simplificar:

$$\begin{aligned} h(t_i; \mathbf{x}_i) &= \lambda \rho (\lambda t_i)^{\rho-1} \exp(\mathbf{x}_i' \beta) = \\ &= \rho t^{\rho-1} \exp(\rho \log \lambda + \mathbf{x}_i' \beta) = \\ &= \rho t^{\rho-1} \exp(\mathbf{x}_i^* \beta^*) \end{aligned} \quad [20]$$

on  $\beta^* = (\rho \log \lambda, \beta')$  i  $\mathbf{x}_i^* = (1, \mathbf{x}_i')$ . Per tant, un únic paràmetre  $\rho$  descriu la funció de risc base i  $\rho \log \lambda$  apareix com un terme independent en l'escala logarítmica. Així, la funció de supervivència d'aquest model es calcularia a partir de la seva definició:

$$\begin{aligned}
 S(t_i; \mathbf{x}_i) &= \exp \left[ - \int_0^{t_i} h(u; \mathbf{x}_i) du \right] = \\
 &= \exp \left[ - \int_0^{t_i} \rho(u)^{\rho-1} \exp(\mathbf{x}_i' \boldsymbol{\beta}^*) du \right] = \\
 &= \exp \left[ - \left( \int_0^{t_i} \rho(u)^{\rho-1} du \right) \exp(\mathbf{x}_i' \boldsymbol{\beta}^*) \right] = \\
 &= \exp \left[ - t_i^\rho \exp(\mathbf{x}_i' \boldsymbol{\beta}^*) \right]
 \end{aligned} \tag{21}$$

La funció de versemblança del model de Weibull

L'expressió general del logaritme de la versemblança pels models paramètrics, assumint censura aleatòria i definint  $y_i = \min(t_i, c_i)$  i el seu indicador de censura  $\delta_i = \text{Ind}(t_i < c_i)$ , és:

$$\log L(\varphi) = ct + \sum_i [\delta_i \log h(y_i)] + \sum_i \log S(y_i)$$

En el cas del model de Weibull,

$$\log h(y_i; \mathbf{x}_i) = \log \left( \rho y_i^{\rho-1} \exp(\mathbf{x}_i' \boldsymbol{\beta}^*) \right)$$

i

$$\log S(y_i; \mathbf{x}_i) = -y_i^\rho \exp(\mathbf{x}_i' \boldsymbol{\beta}^*)$$

Per tant, la funció de log-versemblança ometent el terme constant seria:

$$\begin{aligned}
 \log L(\rho, \boldsymbol{\beta}^*) &= \sum_i \left[ \delta_i \left( \log \rho + \log(y_i^{\rho-1}) + \log \left( \exp(\mathbf{x}_i' \boldsymbol{\beta}^*) \right) \right) \right] - \sum_i y_i^\rho \exp(\mathbf{x}_i' \boldsymbol{\beta}^*) = \tag{22} \\
 &= N \log \rho + (\rho - 1) \sum_i [\delta_i \log y_i] + \sum_i [\delta_i \mathbf{x}_i' \boldsymbol{\beta}^*] - \sum_i y_i^\rho \exp(\mathbf{x}_i' \boldsymbol{\beta}^*)
 \end{aligned}$$

on tenim  $N = \sum_i \delta_i$  observacions no censurades. Els estimadors de màxima versemblança de  $\rho$  i  $\boldsymbol{\beta}^* = (\rho \log \lambda, \boldsymbol{\beta})$ , s'obtenen maximitzant aquesta expressió.

### **Model semiparametric de Cox**

Tot i que especificar una distribució paramètrica coneguda per la funció de risc base permet precisar i simplificar les estimacions dels coeficients  $\beta$ , un atractiu del model de riscos proporcionals és que permet estimar els coeficients de regressió  $\beta$  sense especificar cap model paramètric conegut per la funció de risc base  $h_0(t)$ , que es pot deixar completament arbitrària (no paramètrica) (Allison, 1995). En aquest model semiparamètric, els coeficients de regressió s'estimen a partir d'una funció de versemblança parcial (Cox, 1972). Aquesta manca d'especificitat li dóna robustesa, ja que permet obtenir estimacions no esbiaixades dels coeficients de regressió en una àmplia varietat de situacions de distribució de les dades i, per tant, les seves estimacions s'aproximen estretament a les obtingudes especificant un model paramètric correcte (Ducrocq, 1997).

#### La funció de versemblança parcial

L'estimació per versemblança parcial (PL) (**PL: Partial Likelihood**) comença novament suposant que per cada individu  $i$  hi ha un temps de vida  $t_i$  i un temps de censura  $c_i$ , però el que realment observem és la variable  $y_i = \min(t_i, c_i)$  i el seu indicador de censura  $\delta_i = \text{Ind}(t_i < c_i)$ . Una funció de versemblança ordinària seria el producte de les versemblances per cadascun dels individus de la mostra  $L_i$ . En canvi, la versemblança parcial és el producte de les versemblances dels registres complets de la mostra. Suposant que tenim  $N$  temps de vida no censurats, això s'escriuria com:

$$PL = \prod_i [L_i]^{\delta_i} = \prod_{j=1}^N L_j \quad [23]$$

on  $L_j$  és la versemblança de la *fallida*  $j$ . Ara el que falta saber és com es construeix cada  $L_j$  individual. Per fer-ho, és convenient ordenar totes les observacions (completes i censurades) en ordre ascendent (del més curt al més llarg).

Suposem que  $t_{(1)} < t_{(2)} < \dots < t_{(j)} \dots < t_{(N)}$  són  $N$  temps de vida no censurats diferents ordenats i que en cadascun d'aquests temps  $t_{(j)}$  únicament hi ha un individu  $i$  que sigui eliminat  $\delta_i = 1$  en el temps  $y_i = t_{(j)}$ . Per construir la funció de versemblança parcial ens hem de preguntar la següent qüestió: donat que s'ha produït

una *fallida* en el temps  $t_{(j)}$ , quina és la probabilitat que li hagi succeït a l'individu  $i$  enlloc de a qualsevol dels altres individus? La resposta és el risc per l'individu  $i$  dividit per la suma dels riscos de tots els individus a risc en aquell temps  $t_{(j)}$ . Suposant que en cada moment  $t_{(j)}$  existeixen  $n_j$  individus a risc, que no han estat reformats ni han estat censurats abans de  $t_{(j)}$ , aquesta probabilitat és:

$$L_j = \frac{h(t_{(j)}; \mathbf{x}_i)}{\sum_{k=1}^{n_j} h(t_{(j)}; \mathbf{x}_k)}$$

Tot i que aquesta fórmula té un considerable atractiu intuïtiu, la seva demostració és realment complicada i no es presenta aquí (veure Cox, 1972).

Fins ara la fórmula és completament general sense fer cap assumptió sobre la forma de la funció de risc. Ara, si invoquem el model de riscos proporcionals i el substituïm en la fórmula anterior:

$$L_j = \frac{h_0(t_{(j)}) \exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_{k=1}^{n_j} h_0(t_{(j)}) \exp(\mathbf{x}'_k \boldsymbol{\beta})} = \frac{h_0(t_{(j)}) \exp(\mathbf{x}'_i \boldsymbol{\beta})}{h_0(t_{(j)}) \left[ \sum_{k=1}^{n_j} \exp(\mathbf{x}'_k \boldsymbol{\beta}) \right]} = \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{\left[ \sum_{k=1}^{n_j} \exp(\mathbf{x}'_k \boldsymbol{\beta}) \right]}$$

on  $\mathbf{x}'_k$  és el valor de  $\mathbf{x}'$  per l'individu  $k$ . Això porta a una considerable simplificació perquè la funció inespecífica  $h_0(t_{(j)})$  és comuna a cadascun dels termes de l'expressió. La seva simplificació fa possible estimar els coeficients del vector  $\boldsymbol{\beta}$  sense haver d'especificar una funció de risc base. Cal remarcar que la funció de versemblança parcial només depèn de l'ordre en que es produeixen les fallides i no del seu valor exacte ja que aquest no apareix en la fórmula final (Allison, 1995).

Finalment, una expressió general per la funció de versemblança parcial d'un model de riscos proporcionals seria:

$$PL = \prod_i \left[ \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta})}{\sum_k I_{ik} \exp(\mathbf{x}'_k \boldsymbol{\beta})} \right]^{\delta_i} \quad [24]$$

on l'indicador  $I_{ik} = \text{Ind}(y_i < y_k)$  permet excloure del denominador els individus que ja no estan a risc. Encara que l'expressió sigui un producte per tots els individus de la mostra enlloc d'únicament els  $N$  que han experimentat la *fallida*, els termes

corresponents a individus censurats són efectivament exclosos perquè  $\delta_i = 0$  per aquests casos.

Una vegada construïda la funció de versemblança parcial, aquesta es pot maximitzar respecte a  $\beta$  igual que en el cas de la versemblança completa. Com és habitual, és preferible maximitzar el logaritme de la versemblança:

$$\log PL = \sum_i \left( \delta_i \left[ \mathbf{x}_i' \beta - \log \left( \sum_k I_{ik} \exp(\mathbf{x}_k' \beta) \right) \right] \right) \quad [25]$$

La majoria de programes de versemblança parcial usen alguna versió de l'algorisme de Newton-Raphson per maximitzar aquesta funció respecte a  $\beta$ .

Tot i que aquesta expressió permet coincidències entre una *fallida* i un o més temps de censura, no és vàlida quan hi ha temps de *fallida* coincidents. Quan això succeeix, és a dir, en un moment  $y_i$  es produeixen un nombre variable de fallides  $d_i$ , la versemblança parcial ha estat formulada amb diferents aproximacions. Quan hi ha poques coincidències, l'aproximació més utilitzada és:

$$PL = \prod_i \left[ \frac{\exp(\mathbf{x}_i' \beta)}{\left( \sum_k I_{ik} \exp(\mathbf{x}_k' \beta) \right)^{d_i}} \right]^{\delta_i} \quad [26]$$

atribuïda a Peto (1972) i Breslow (1974). Aquesta és l'aproximació utilitzada en la majoria dels paquets estadístics i, més específicament, en el Survival Kit (Ducrocq i Sölkner, 1998). Quan  $d_i$  és gran, existeixen altres aproximacions (veure Klein i Moeschberger, 1997).

### La funció de supervivència base

Una vegada estimats els paràmetres de  $\beta$ , pot ser interessant fer una estimació no paramètrica de la distribució base. Assumint que la funció de risc base és constant entre fallides, un estimador de la funció de risc base al temps  $t$  entre les fallides observades  $t_{(j-1)}$  i  $t_{(j)}$  és:



$$\hat{h}_0(t) = \hat{h}_{(j)} = \frac{d_j}{(t_{(j)} - t_{(j-1)}) \left[ \sum_{k=1}^{n_j} \exp(\mathbf{x}_k' \boldsymbol{\beta}) \right]} \quad [27]$$

on  $d_j$  és el nombre de fallides observades a  $t_{(j)}$ . Tsiatis (1981) va suggerir usar aquests valors per estimar la funció de risc acumulat base de la següent manera:

$$\hat{H}_0(t) = \int_0^t \hat{h}_0(u) du = \sum_j \left( I_{jk} \frac{d_j}{\left[ \sum_{k=1}^{n_j} \exp(\mathbf{x}_k' \boldsymbol{\beta}) \right]} \right) \quad [28]$$

on  $I_{jk} = \text{Ind}(t_{(j)} < y_k)$  permet excloure del denominador els individus que ja han experimentat la *fallida* i, per tant, no formen part dels individus a risc. Finalment, l'estimador corresponent de la funció de supervivència base és:

$$\hat{S}_0(t) = \exp[-\hat{H}_0(t)] \quad [29]$$

### La hipòtesi de proporcionalitat

Tant si es tracta d'un model paramètric com d'un model semiparamètric, el nom de model de riscos proporcionals prové de l'assumpció de la següent propietat: si  $h(t; \mathbf{x}_A)$  i  $h(t; \mathbf{x}_B)$  són les funcions de risc de dos grups d'individus amb valors diferents  $\mathbf{x}_A$  i  $\mathbf{x}_B$  del vector de variables explicatives, la raó d'aquests dos riscos (**HR = Hazard Ratio**) és:

$$HR = \frac{h(t; \mathbf{x}_A)}{h(t; \mathbf{x}_B)} = \frac{h_0(t) \exp(\mathbf{x}_A' \boldsymbol{\beta})}{h_0(t) \exp(\mathbf{x}_B' \boldsymbol{\beta})} = \exp((\mathbf{x}_A' - \mathbf{x}_B') \boldsymbol{\beta}) \quad [30]$$

Per tant, la raó de risc és una constant de proporcionalitat que depèn dels valors de les variables explicatives, però no depèn del temps de vida. Per exemple, si la raó de risc és igual a 2, els individus del grup A tenen el doble de risc de *fallida* que els individus del grup B en qualsevol moment de la vida. Per tant, en el fons el que estem assumint és que la funció de risc base  $h_0(t)$  és la mateixa per tots els individus i que l'efecte de les variables explicatives que multiplica al valor de la funció de risc base és el mateix en cada temps de vida.

El problema és que aquesta hipòtesi de proporcionalitat no sempre és assumible en totes les variables explicatives. Quan la hipòtesi de proporcionalitat no es compleix, llavors el model de riscos proporcionals no és apropiat. La no proporcionalitat és la interacció de l'efecte de les variables amb el temps. Si no tenim en compte aquesta interacció, el coeficient estimat és com l'efecte mitjà de la variable al llarg dels temps (Allison, 1995). En el model lineal normal i en el model AFTM també s'assumeixen hipòtesis semblants. Per exemple, per establir la dependència dels pesos al naixement en el sexe en un model lineal normal assumim que la distribució dels pesos al naixement és normal amb la mateixa variància en els mascles i en les femelles i que la diferència de pesos al naixement entre sexes (l'efecte additiu del sexe) és el mateix independentment de si la mesurem en els animals amb pesos petits (el primer quartil de cada sexe), mitjans (la mediana de cada sexe) o grans (el tercer quartil de cada sexe).

Malgrat això, a vegades aquesta interacció de l'efecte de les variables amb el temps ens interessa teòricament o és tan forta que s'ha d'estudiar. Quan la hipòtesi de proporcionalitat no es pot assumir en el model de regressió lineal (per exemple, les variàncies dels pesos al naixement són heterogènies entre sexes) o el model AFT, la solució és transformar la variable resposta, fer una anàlisi amb pesos o fer anàlisis separades per cadascuna de les categories heterogènies. En canvi, en el model PHM, les solucions a la no proporcionalitat són: (i) l'estratificació, que consisteix en considerar distribucions base diferents per cadascuna de les categories de les variables heterogènies (o no proporcionals), però realitzant una estimació conjunta dels coeficients de regressió de les variables homogènies (o proporcionals) o (ii) ampliar el model PHM per incloure variables dependents de la variable resposta (o sigui, del temps).

El model de riscos proporcionals estratificat consisteix en analitzar un model que inclogui les variables explicatives proporcionals  $\mathbf{x}$  considerant distribucions base diferents per cadascuna de les categories de la variable no proporcional (Kleinbaum, 1996). Per tant, la funció de risc del model de riscos proporcionals estratificat per les  $g = 1, 2, \dots, k$  categories de la variable no proporcional és:

$$h_g(t_i, \mathbf{x}_i) = h_{0g}(t_i) \exp(\mathbf{x}_i' \boldsymbol{\beta}) \quad [31]$$

En aquest model permetem una funció de risc base diferent per cada categoria  $g$ . Per exemple, si l'efecte del sexe no fóra proporcional, considerariem una distribució base pels mascles i una de diferent per les femelles. L'avantatge del model estratificat és

que no s'ha d'especificar cap forma particular per evolució al llarg del temps de vida de l'efecte de la variable no proporcional, donant més robustesa a les estimacions de l'efecte de les variables proporcionals incloses en el model  $\exp(\mathbf{x}_i' \boldsymbol{\beta})$ . L'inconvenient és que no permet estimar l'efecte de la variable no proporcional ni la seva interacció amb el temps de vida (Allison, 1995). L'estratificació és molt útil quan les variables no proporcionals són categòriques i sense interès directe (per exemple, l'efecte de l'explotació).

Per altra banda, l'ampliació del model de riscos proporcionals amb variables dependents del temps consisteix en incloure termes d'interacció  $\mathbf{x}_i(t)$  entre l'efecte de la variable no proporcional  $\mathbf{x}_i$  i diferents funcions del temps de vida  $g_i(t)$  (Kleinbaum, 1996). Per exemple, si volguéssim considerar l'efecte del pes al naixement ( $PN$ ) com a dependent del temps de vida, definiríem la variable  $PN(t) = g_{PN}(t) PN$ . Així, el model de riscos proporcionals ampliat general és:

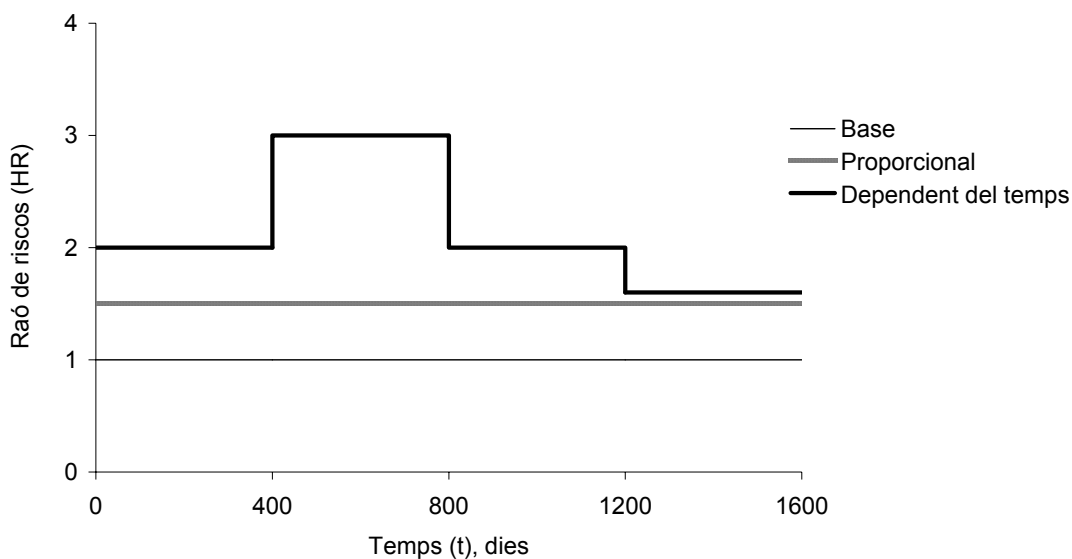
$$h(t, \mathbf{x}_i(t)) = h_0(t) \exp(\mathbf{x}_i(t)' \boldsymbol{\beta}) \quad [32]$$

on el vector  $\boldsymbol{\beta}$  està format pels efectes dels termes d'interacció  $\mathbf{x}_i(t)$ . L'avantatge de l'ampliació del model de riscos proporcionals amb variables dependents del temps de vida és que permet estimar l'evolució de l'efecte d'una variable al llarg del temps. L'ús de variables dependents de la variable resposta permet obtenir estimacions més precises de l'efecte d'aquestes variables no proporcionals (Allison, 1995).

El problema de l'ús de variables dependents del temps és l'elecció de la funció del temps  $g_i(t)$  que interacciona amb cadascuna de les variables, ja que poden ser molt diverses ( $t$ ,  $\log t$ , funcions constants a intervals, etc.). Això suposa una notable complicació computacional. En el Survival Kit (Ducrocq i Sölkner, 1998), per facilitar el càlcul de la versemblança, únicament es tenen en compte variables dependents del temps constants a intervals. Aquestes assumeixen que la premissa de riscos proporcionals es manté dins d'uns períodes de temps i deixa de complir-se quan es produeix un punt de tall, originant-se un nou període de riscos proporcionals fins al nou punt de tall, i així successivament (Figura 1). En funció de la naturalesa del punt de tall, les variables dependents del temps constants a intervals poden ser classificades en externes i internes (Ducrocq, 2000). Les **variables externes** no estan influïdes per cadascun dels individus en estudi. Poden ser variables definides amb punts de tall del temps coneguts abans de començar l'estudi. Per exemple, quan l'efecte d'una variable és diferent abans i després d'una determinada edat. Però els punts de tall també

poden no ser coneguts com, per exemple, la temperatura ambiental que varia amb el temps. Les **variables internes** són el resultat d'un procés estocàstic generat per l'individu en estudi. Per exemple, quan l'efecte d'una variable canvia abans i després de cada part, el punt de tall es la data de part que té una variabilitat entre individus.

**Figura 1.** Variables proporcionals versus variables dependents del temps.



### **Validació de la hipòtesi de proporcionalitat**

Existeixen diferents mètodes per comprovar si les variables incloses en el model compleixen la hipòtesi de proporcionalitat: el mètode d'estratificació i comparació gràfica i el mètode de les variables dependents del temps.

#### Mètode d'estratificació i comparació gràfica

El mètode d'estratificació té una sèrie de passos que repetirem per cada variable a contrastar (Kleinbaum, 1996). Si la variable a contrastar és continua, primer haurem de categoritzar les dades en  $k$  categories diferents. Un cop definides les  $k$  categories de la variable a contrastar, estimarem les funcions de supervivència empíriques per cadascuna de les  $g = 1, 2, \dots, k$  categories. Això es pot realitzar utilitzant diferents mètodes: i) el mètode de Kaplan-Meier estratificant per la variable testada  $\hat{S}_{KMg}(t)$  (equació 9); o ii) un model de Cox estratificat per la variable testada corregint per les

altres variables d'elecció assumides proporcionals  $\hat{S}_{0g}(t)$  (equació 29). Aquest mètode és preferible perquè corregeix l'efecte de les altres variables (Kleinbaum, 1996). Un cop tenim les supervivències empíriques observades podem comprovar la hipòtesi de proporcionalitat de diferents maneres (veure Kleinbaum, 1996). El mètode dels logaritmes consisteix en calcular el logaritme del logaritme negatiu de les supervivències observades dels  $k$  estrats,  $\log(-\log \hat{S}_g(t))$ , i representar-los gràficament respecte al logaritme del temps de *fallida*,  $\log t$ . Si la variable és proporcional, existirà paral·lelisme entre categories. Una manca de paral·lelisme important indicaria manca de proporcionalitat (Allison, 1995). Aquest és el mètode dels logaritmes utilitzat anteriorment per validar la distribució de Weibull. Aquests mètodes gràfics són útils quan tenim moltes dades. Quan en tenim poques, tot i que s'observi una manca de proporcionalitat, no podem assegurar si aquesta diferència és important o no.

#### Mètode de les variables dependents del temps

Aquest mètode consisteix en utilitzar models que incloguin variables dependents del temps ampliant el model de riscos proporcionals amb termes d'interacció  $\mathbf{x}_i(t)$  entre la variable a contrastar  $\mathbf{x}_i$  i una funció del temps  $g_i(t)$ . Si l'efecte del terme d'interacció amb el temps és significatiu, llavors es rebutja la hipòtesi de proporcionalitat en la variable testada. El problema d'aquest mètode és novament l'elecció de la funció del temps  $g_i(t)$  que interacciona amb cada variable independent a contrastar. Diferents eleccions poden portar a conclusions diferents sobre la proporcionalitat (Allison, 1995). Un altre problema és que si les mostres són grans, qualsevol test tendirà a rebutjar la hipòtesi de proporcionalitat (Klein i Moeschberger, 1997).

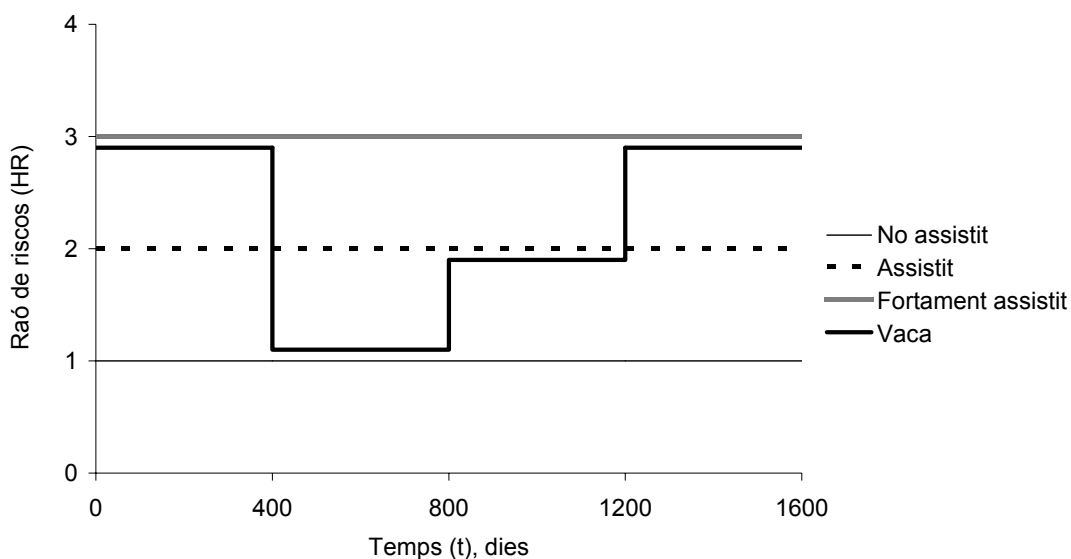
#### **Altres tipus de variables dependents del temps**

Les variables dependents del temps presentades fins ara consideraven que un individu pertanyia a un mateix nivell de la variable durant tota la seva vida i que l'efecte d'aquest nivell sobre el risc interaccionava amb el temps de vida. No obstant, el model de riscos proporcional també permet incloure variables que tenen mesures repetides al llarg de la vida de l'individu. En aquest cas, les expressions generals que proporcionen la versemblança romanen inalterades, però s'ha de tenir en compte els canvis en els

valors de les variables que es produeixen al llarg del temps (Ducrocq, 2000). Podem distingir dos tipus de variables amb mesures repetides:

Variables amb mesures repetides proporcionals: Un individu pot canviar de nivell cada vegada que es mesura una variable al llarg de la seva vida però l'efecte dins un mateix nivell és proporcional. Per exemple, una vaca pareix vedells amb diferent dificultat a cada part (el valor de la dificultat de part canvia a cada part) però l'efecte de tenir, per exemple, un part fortament assistit és sempre el mateix respecte a un part no assistit (veure figura 2). En aquest cas, si un individu es manté en el mateix nivell en totes les mesures repetides, el seu efecte serà proporcional al llarg de la vida i l'efecte només dependrà del temps quan els individus canvien de nivell al llarg de la vida. Per exemple, en la figura 2 veiem com canvia al llarg del temps l'efecte d'una vaca que tingui un primer part fortament assistit, un segon part no assistit, un tercer part assistit i finalment un part fortament assistit.

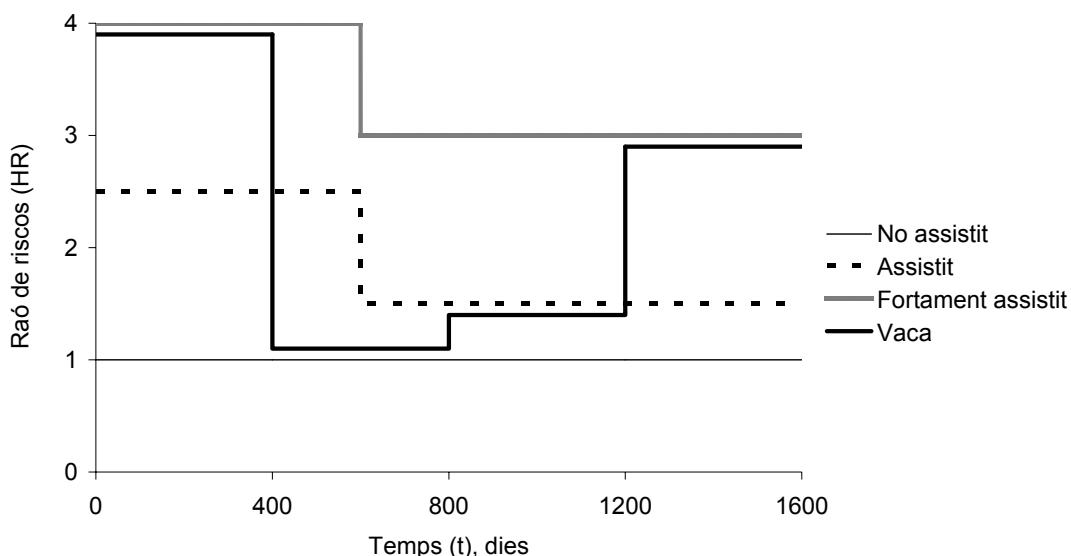
**Figura 2.** Exemple de variable amb mesures repetides proporcionals.



Variables amb mesures repetides que interaccionen amb el temps: Un individu pot canviar de nivell cada vegada que es mesura una variable al llarg de la seva vida i, a més, l'efecte dins un nivell interacciona amb el temps (veure figura 3). Per exemple, l'efecte de tenir un part fortament assistit és més gran al principi de la vida productiva que al final respecte a un part no assistit. En aquest cas, encara que un individu es

mantingui en el mateix nivell en totes les mesures repetides, el seu efecte variarà al llarg de la vida. Novament, la figura 2 mostra com canvia al llarg del temps l'efecte d'una vaca que tingui un primer part fortament assistit, un segon part no assistit, un tercer part assistit i finalment un part fortament assistit.

**Figura 3.** Exemple de variable amb mesures repetides que interacciona amb el temps.



### Raó de riscos

La mesura de l'efecte d'una variable és la raó de riscos (HR). Aquesta generalment és calculada designant un grup per ser la categoria de referència  $\mathbf{x}_r$  i mesurant totes les altres categories en relació amb ella. El risc de la categoria de referència és igual al risc base  $h(t; \mathbf{x}_r) = h_0(t)$ . Així la raó de riscos s'escriu com:

$$HR = \frac{h(t; \mathbf{x}_A)}{h(t; \mathbf{x}_r)} = \frac{h_0(t) \exp(\mathbf{x}_A' \boldsymbol{\beta})}{h_0(t)} = \exp(\mathbf{x}_A' \boldsymbol{\beta}) \quad [33]$$

Per això, la categoria de referència té un  $HR = 1$ . Si el  $HR < 1$ , els individus del grup  $A$  tenen menor risc que els del grup de referència. Inversament, si el  $HR > 1$ , els individus del grup  $A$  tenen major risc que els del grup de referència (Allore i col·l., 2001).

En el cas del model de riscos proporcionals ampliat amb variables dependents del temps  $\mathbf{x}(t)$ , la raó de riscos no és constant i varia en funció del temps. Per tant, la raó de riscos s'haurà de calcular en diferents moments del temps (Kleinbaum, 1996):

$$HR(t) = \frac{h(t; \mathbf{x}_A(t))}{h(t; \mathbf{x}_r)} = \frac{h_0(t) \exp(\mathbf{x}_A(t)' \boldsymbol{\beta})}{h_0(t)} = \exp(\mathbf{x}_A(t)' \boldsymbol{\beta}) \quad [34]$$

### **Funcions de supervivència esperades**

La raó de riscos pot ser difícil d'interpretar i és més interessant veure com canvia la mitjana o la mediana del temps de vida. Aquests canvis depenen de la raó de riscos entre dos grups però també de la funció de distribució base. En aquest sentit, és interessant obtenir la funció de supervivència esperada de cada grup. Un cop estimada la funció de supervivència base  $S_0(t)$ , ja sigui paramètricament o no paramètricament (veure equació 29), la funció de supervivència esperada  $S(t, \mathbf{x}_A)$  d'un individu del grup  $A$  és:

$$S(t, \mathbf{x}_A) = [S_0(t)]^{\exp(\mathbf{x}_A' \boldsymbol{\beta})} = [S_0(t)]^{HR_A} \quad [35]$$

aquesta s'obté elevant la funció de supervivència base  $S_0(t)$  a la potència de la raó de risc  $HR_A$  del grup  $A$ .

### **Models de riscos proporcionals mixtes**

Els models de riscos proporcionals, tant paramètrics com semiparamètrics, poden incloure efectes aleatoris a l'igual que els models mixtes lineals usats arreu del món per les avaluacions genètiques. Els models de riscos proporcionals mixtes clàssicament s'anomenen models de fragilitat (Frailty models) (Vaupel i col·l., 1979; Clayton i Cuzick, 1985; Aalen, 1994). El terme de fragilitat,  $v$ , es defineix com una quantitat aleatòria no observada que afecta multiplicativament el risc dels individus o grup d'individus (Ducrocq, 1997). Així, els individus amb un major valor per al terme de fragilitat és més probable que experimentin la *fallida* abans que els individus amb un valor petit. Dit d'una altra manera, els individus més fràgils solen ser eliminats abans que els menys fràgils. La forma habitual d'incloure el terme de fragilitat  $v_i$  és a través d'un factor de naturalesa aleatòria que actua multiplicativament sobre el risc de *fallida*.



$$h(t_i; \mathbf{x}_i) = h_0(t_i) v_i \exp(\mathbf{x}_i' \boldsymbol{\beta})$$

Quan el terme  $v_i$  és definit per cada individu  $i$ , el component de fragilitat extrau part de la variació no observada entre individus (Vaupel i col·l., 1979; Hougaard, 1986a; Hougaard, 1986b; Aalen, 1994) i també permet una correcció de la possible discrepància entre la variància vertadera de les observacions i l'especificada pel model (Ducrocq, 1997). Aquesta variació extra s'anomena sobredispersió (Louis, 1991; Tempelman i Gianola, 1994). Quan  $v_q$  és definida per un grup d'individus (per exemple, totes les filles d'un toro  $q$ ), descriu les característiques compartides no observables (genètiques en aquest cas) que actuen sobre el risc de cada membre del grup (Clayton i Cuzick, 1985; Anderson i col·l., 1992; Klein, 1992; Klein i col·l., 1992). En tots els casos, la simple transformació  $\mathbf{a} = \log \mathbf{v}$  permet incloure el terme de fragilitat en la part exponencial del model de riscos proporcionals. Si  $\mathbf{z}_i$  representa el vector d'incidència per l'individu  $i$  dels efectes aleatoris  $\mathbf{a}$ , el model de riscos proporcionals mixt pot ser escrit com:

$$h(t_i; \mathbf{x}_i; \mathbf{z}_i) = h_0(t_i) \exp(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a}) \quad [36]$$

on la funció de risc base  $h_0(t_i)$  pot tenir una forma paramètrica o no paramètrica. L'expressió pot ser ampliada per incloure estratificació o variables dependents del temps fixes o aleatòries (Ducrocq, 1997). Les possibles distribucions associades al terme de fragilitat i les aproximacions proposades per estimar els seus paràmetres estan descrites a la bibliografia (Ducrocq i Casella, 1996; Klein i Moeschberger, 1997; Ducrocq, 2000).

### **Models de Weibull mixtes**

En aquesta secció es presenta una aproximació general per l'anàlisi de models de fragilitat, pensant en situacions típiques de millora genètica. El desenvolupament es realitza per un model animal de Weibull simple amb una distribució a priori log-normal pel terme de fragilitat. Per simplificar, només es considera un efecte aleatori  $\mathbf{a}$  en el model. No obstant, es poden fer generalitzacions correctes a altres distribucions a priori (Log-Gamma) i a altres models (amb estratificació i variables dependents del temps, models de Cox, ...) (Ducrocq i Casella, 1996; Ducrocq, 2000).

**Distribució a priori del terme de fragilitat**

En genètica quantitativa, degut al model poligènic infinitesimal usualment assumit, és natural considerar que la distribució de l'efecte additiu de l'animal  $\mathbf{a}$  és normal multivariada:

$$\mathbf{a} \sim NMV(0, \mathbf{A}\sigma_a^2)$$

on  $\mathbf{A}$  és la matriu de parentescos (relacions additives) entre animals. Per tant, la densitat de  $\mathbf{a}$  és:

$$\pi_0(\mathbf{a} | \sigma_a^2) = \frac{1}{(2\pi\sigma_a^2)^{N/2} |\mathbf{A}|^{1/2}} \exp\left(-\frac{1}{2\sigma_a^2} \mathbf{a}' \mathbf{A}^{-1} \mathbf{a}\right) \quad [37]$$

on  $\sigma_a^2$  és la variància additiva de l'animal.

**La funció de versemblança del model de Weibull**

L'expressió general del logaritme de la funció de versemblança pels models paramètrics, assumint censura aleatòria i definint  $y_i = \min(t_i, c_i)$  i el seu indicador de censura  $\delta_i = \text{Ind}(t_i < c_i)$ , és:

$$\log L(\mathbf{y} | \boldsymbol{\theta}, \rho) = ct + \sum_i [\delta_i \log h(y_i)] + \sum_i \log S(y_i)$$

En el cas del model de Weibull mixt,

$$\log h(y_i; \mathbf{w}_i) = \log\left(\rho y_i^{\rho-1} \exp(\mathbf{w}_i' \boldsymbol{\theta})\right)$$

i

$$\log S(y_i; \mathbf{w}_i) = -y_i^\rho \exp(\mathbf{w}_i' \boldsymbol{\theta})$$

on la matriu d'incidència  $\mathbf{w}_i' = \begin{pmatrix} \mathbf{x}_i' & \mathbf{z}_i' \end{pmatrix}$  i el vector de paràmetres  $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{a} \end{pmatrix}$  inclouen efectes fixes i aleatoris. Per tant, la funció de log-versemblança, ometent el terme constant, seria:

$$\log L(\mathbf{y} | \boldsymbol{\theta}, \rho) = N \log \rho + (\rho - 1) \sum_i [\delta_i \log y_i] + \sum_i [\delta_i \mathbf{w}_i' \boldsymbol{\theta}] - \sum_i y_i^\rho \exp(\mathbf{w}_i' \boldsymbol{\theta}) \quad [38]$$

on tenim N observacions no censurades.

**Densitat posterior conjunta**

Aplicant el teorema de Bayes, tenim que:

$$p(\boldsymbol{\theta}, \rho, \sigma_a^2 | \mathbf{y}) \propto L(\mathbf{y} | \boldsymbol{\theta}, \rho) \pi_0(\mathbf{a} | \sigma_a^2) \pi_0(\boldsymbol{\beta}, \rho) \pi_0(\sigma_a^2) \quad [39]$$

on assumim un a priori pla per  $\sigma_a^2$  i per  $(\boldsymbol{\beta}, \rho)$  (Ducrocq, 2000).

***Inferència sobre  $\boldsymbol{\theta}$  i  $\rho$***

Agafant logaritmes als dos costats de l'equació anterior tenim que:

$$\log p(\boldsymbol{\theta}, \rho, \sigma_a^2 | \mathbf{y}) = ct + \log L(\mathbf{y} | \boldsymbol{\theta}, \rho) + \log \pi_0(\mathbf{a} | \sigma_a^2)$$

Si assumim que  $\sigma_a^2$  és coneguda, el logaritme de la densitat posterior conjunta de  $(\boldsymbol{\theta}, \rho)$  donada  $\sigma_a^2$  és (Ducrocq, 2000):

$$\begin{aligned} \log p(\boldsymbol{\theta}, \rho | \mathbf{y}, \sigma_a^2) = & N \log \rho + (\rho - 1) \sum_i [\delta_i \log y_i] + \sum_i [\delta_i \mathbf{w}_i' \boldsymbol{\theta}] - \\ & - \sum_i y_i^\rho \exp(\mathbf{w}_i' \boldsymbol{\theta}) - \frac{1}{2\sigma_a^2} \mathbf{a}' \mathbf{A}^{-1} \mathbf{a} \end{aligned} \quad [40]$$

***Inferència sobre  $\sigma_a^2$***

La inferència sobre el paràmetre de dispersió  $\sigma_a^2$  està basada en la seva distribució marginal posterior, després d'integrar els paràmetres  $\boldsymbol{\theta}$  i  $\rho$  de la densitat posterior conjunta (Berger, 1985; Robert, 1992):

$$p(\sigma_a^2 | \mathbf{y}) = \iint p(\boldsymbol{\theta}, \rho, \sigma_a^2 | \mathbf{y}) \delta \boldsymbol{\theta} \delta \rho \quad [41]$$

Excepte en casos trivials, aquesta integració no es pot realitzar algebraicament (Ducrocq, 2000). Per obtenir la distribució marginal posterior del paràmetre de dispersió  $\sigma_a^2$ , es pot:

- a) simular mostres aleatòries a partir d'aquesta distribució (Clayton, 1991; Dellaportas i Smith, 1993; Korsgaard, 1996; Korsgaard i col·l., 1998).
- b) calcular la integral numèricament (Smith i col·l., 1985).

c) trobar una aproximació.

Ducrocq i Casella (1996) van triar la tercera alternativa, usant una tècnica coneguda com integració Laplaciana (Achcar i Bolfarine, 1986; Tierney i Kardane, 1986; Tierney i col·l., 1989; Tempelman i Gianola, 1993; Goutis i Casella, 1996). Aquesta és l'aproximació utilitzada en el paquet estadístic Survival Kit (Ducrocq i Sölkner, 1998) per obtenir una estimació modal de  $\sigma_a^2$ .

### **Heretabilitat i fiabilitat del temps de vida**

#### ***Escala original***

Donada la naturalesa no lineal dels models usats en l'anàlisi de supervivència, no és obvi quina és la definició adequada dels paràmetres genètics, i més en concret, de l'heretabilitat, existint varies formes d'estimar-la. Una d'elles consisteix a obtenir-la com a quocient de variàncies (Ducrocq i col·l., 1988a; Ducrocq i Casella, 1996; Ducrocq, 1999; Ducrocq, 2000; Korsgaard i col·l., 1999; Korsgaard i col·l., 2000; Korsgaard i col·l., 2002). No obstant, aquestes formes no permeten predir la precisió obtinguda en les avaluacions genètiques com una funció de les dades disponibles (per exemple, com una funció del nombre de filles d'un progenitor). Per poder calcular un "coeficient de fiabilitat" aproximat  $F$ , finalment s'ha imposat una expressió denominada heretabilitat efectiva que, donat l'estimador de la variància pare  $\hat{\sigma}_s^2$ , és:

$$h^2 = \frac{4\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + 1} \quad [42]$$

Yazdi i col·l. (2002) la van obtenir igualant la fiabilitat vertadera, obtinguda a partir del càlcul de la variància de l'error de predicció dels valors genètics i la fiabilitat aproximada. Aquesta expressió va ser desenvolupada assumint una distribució Weibull i un model pare, que és el més utilitzat usualment i per al que s'han derivat gran part de les fórmules d'estimació dels paràmetres d'interès. La seva extensió al model animal o a altres models que inclouen més factors aleatoris no és complicada.

Aquesta heretabilitat correspon a l'heretabilitat del caràcter en absència de censura. No obstant, quan es prenen les decisions de selecció, les observacions no censurades representen una proporció baixa de totes les observacions. Per tant, l'heretabilitat efectiva es veu reduïda. Yazdi i col·l. (2002) proposaren una heretabilitat

equivalent per a models pare quan només una proporció  $p$  dels animals tenen registres no censurats:

$$h_e^2(t) = \frac{4\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + \frac{1}{p}} \quad [43]$$

Aquestes heretabilitats equivalents són útils per calcular les fiabilitats aproximades dels valors genètics estimats per als pares. Per exemple, si un pare amb  $N$  filles té com a mitjana  $N_u$  filles amb registre no censurat després de  $t$  dies de vida productiva, la fiabilitat aproximada del valor genètic del pare basada en les seves filles és:

$$F = \frac{N_u h^2}{(N_u - 1)h^2 + 4} = \frac{N h_e^2(t)}{(N - 1)h_e^2(t) + 4} \quad [44]$$

Amb aquesta fórmula es pot predir la resposta a la selecció usant  $R = i f \sigma_s$ , on  $i$  és la intensitat de selecció,  $f$  és l'arrel quadrada de la fiabilitat  $F$  i  $\sigma_s$  és l'arrel quadrada de la variància del pare (Yazdi i col·l., 2002).

### **Escala binària**

L'escala binària és un cas especial d'una mesura discreta de la supervivència amb només dos possibles valors: 0 o 1, depenent de si l'animal segueix viu en un temps donat (per exemple, al deslletament). Quan el temps de vida segueix una distribució de Weibull, l'heretabilitat en l'escala binària  $h_{bin}^2$  pot ser aproximada com (Yazdi i col·l., 2002):

$$h_{bin}^2 \approx \frac{4\sigma_s^2}{\sigma_s^2 + \frac{p}{[\ln(1-p)]^2(1-p)}} \quad [45]$$

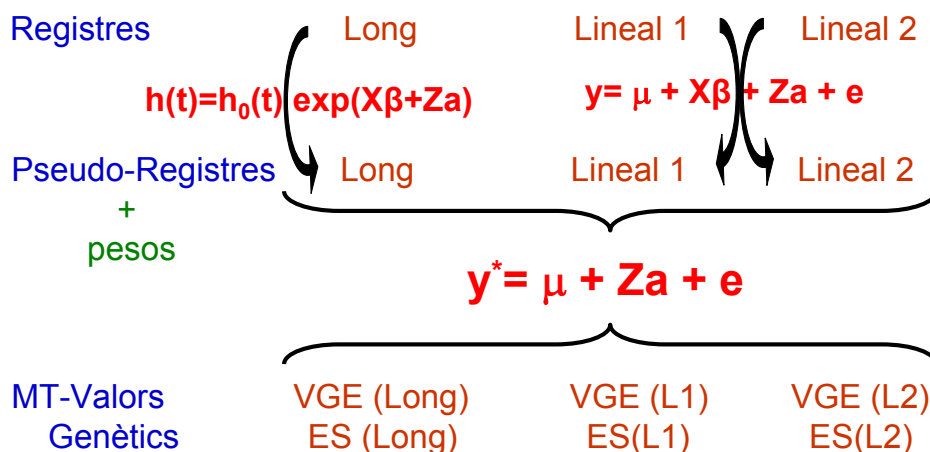
A partir d'aquesta fórmula, la fiabilitat en l'escala binària es prediu usant la teoria d'índexs de selecció com:

$$F_{bin} \approx \frac{n}{n + \frac{(4 - h_{bin}^2)}{h_{bin}^2}} \quad [46]$$

**Anàlisi multicaràcter**

Fins ara hem proposat l'avaluació genètica de la supervivència amb una aproximació unicaràcter. No obstant, una aproximació multicaràcter permetria obtenir una major precisió de les avaluacions gràcies a un increment en la quantitat d'informació utilitzada, una millor estructura de dades deguda a una millor connexió i una correcció del biaix per la selecció en caràcters correlacionats (Ducrocq i col·l., 2001). El problema per poder combinar dades de longevitat amb informació indirecta provenint d'altres caràcters és que els models d'anàlisi univariants actuals pels diferents caràcters són tan diferents (han de permetre censura, no linealitat i no normalitat del residu) que una aproximació multicaràcter exacta, encara que sigui conceptualment possible, no és factible rutinàriament (Ducrocq i col·l., 2001). Malgrat les grans i ràpides millores en potència de computació, les consideracions computacionals són encara un factor limitant. Per evitar-ho, una aproximació menys demandant amb dos passos va ser proposada per Ducrocq i col·l. (2001).

**Figura 4.** Esquema general de l'aproximació multicaràcter amb dos passos.



En el primer pas, es fan anàlisis univariats amb els models actuals per estimar les variàncies genètiques i residuals i per calcular els pseudo-registres i els pesos associats de tots els caràcters. Aquests pseudo-registres són registres virtuals lliures d'efectes ambientals que poden ser usats en un model animal BLUP per obtenir els mateixos valors additius que en el model univariant actual (possiblement no lineal). En el segon pas, es combinen aquests pseudo-registres en un model animal multicaràcter

per obtenir estimacions de les correlacions genètiques i residuals mitjançant un AI-REML. Aquestes correlacions permeten estimar mitjançant un BLUP multicaràcter els valors genètics que combinen la informació directa i indirecta sobre longevitat.

### **Derivació dels pseudo-registres i pesos per longevitat**

La derivació dels pseudo-registres i pesos per longevitat va ser proposada per Ducrocq (2001a). La derivació es basa en el principi que l'ús dels pseudo-registres  $y_i^*$  (amb els seus pesos  $w_i$ ) en l'avaluació BLUP unicaràcter amb un model animal ha de donar estimacions del valor additiu  $a_i$  d'un determinat animal tan semblants com sigui possible a les estimacions basades en el model de riscos proporcionals mixt. Per a l'estimació del valor additiu  $a_i$  d'un determinat animal amb un model animal BLUP lineal:

$$y_i^* = \mu + a_i + e_i \quad [47]$$

amb  $\text{var}(e_i) = \left(\frac{1}{w_i}\right)\sigma_e^2$ , utilitzem la següent equació:

$$\left[ \mathbf{W} + \frac{1}{\sigma_a^2} \mathbf{A}^{-1} \right] \mathbf{a} = \mathbf{W} \mathbf{y}^*$$

on considerem que  $\mathbf{W} = \text{diag}\{w_i\}$  i que  $\sigma_e^2 = 1$  (Ducrocq, 2001a). L'equació de l'animal  $i$  seria:

$$w_i a_i + \frac{1}{\sigma_a^2} (\mathbf{A}^{-1} \mathbf{a})_i = w_i y_i^* \quad [48]$$

En canvi, per a l'estimació del valor genètic additiu amb un model de riscos proporcionals animal mixt, assumint que  $\sigma_a^2$  és coneguda, utilitzaríem la densitat posterior conjunta de  $(\boldsymbol{\theta}, \rho)$  donada  $\sigma_a^2$  que, aplicant el teorema de Bayes, és:

$$p(\boldsymbol{\theta}, \rho, \sigma_a^2 | \mathbf{y}) \propto L(\mathbf{y} | \boldsymbol{\theta}, \rho) \pi_0(\mathbf{a} | \sigma_a^2)$$

on la funció de versemblança és:

$$L(\mathbf{y} | \boldsymbol{\theta}, \rho) \propto \prod_i [h(y_i)]^{\delta_i} S(y_i) =$$

$$= \prod_i \left[ h_0(y_i) \exp(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a}) \right]^{\delta_i} [S_0(y_i)]^{\exp(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a})}$$

on  $\delta_i$  és l'indicador de censura ( $\delta_i = 0$  si l'animal  $i$  és censurat i  $\delta_i = 1$  si no). Així, el logaritme de la densitat posterior conjunta de  $(\boldsymbol{\theta}, \rho)$  donada  $\sigma_a^2$  és:

$$\log p(\boldsymbol{\theta}, \rho | \mathbf{y}, \sigma_a^2) = ct + \sum_i [\delta_i \log h_0(y_i)] + \sum_i [\delta_i (\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a})] -$$

$$- \sum_i H_0(y_i) \exp(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a}) - \frac{1}{2\sigma_a^2} \mathbf{a}' \mathbf{A}^{-1} \mathbf{a}$$

on  $H_0(y_i) = -\log(S_0(y_i))$  és el risc acumulat base. Igualant a 0 la primera derivada d'aquesta equació respecte al valor additiu  $a_i$  de l'animal  $i$ , obtenim que:

$$\frac{\delta}{\delta a_i} \left( \sum_i [\delta_i (\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a})] - \sum_i H_0(y_i) \exp(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a}) - \frac{1}{2\sigma_a^2} \mathbf{a}' \mathbf{A}^{-1} \mathbf{a} \right) = 0$$

o sigui:

$$\delta_i - H_0(y_i) \exp(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a}) - \frac{1}{\sigma_a^2} (\mathbf{A}^{-1} \mathbf{a})_i = 0$$

Si definim el pes com a  $w_i = H(y_i; \mathbf{x}_i; \mathbf{z}_i) = H_0(y_i) \exp(\mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_i' \mathbf{a})$ , l'equació es reescriu com:

$$w_i + \frac{1}{\sigma_a^2} (\mathbf{A}^{-1} \mathbf{a})_i = \delta_i$$

Com que  $1 = a_i + (1 - a_i)$ , l'equació esdevé:

$$w_i [a_i + (1 - a_i)] + \frac{1}{\sigma_a^2} (\mathbf{A}^{-1} \mathbf{a})_i = \delta_i$$

i, per tant,

$$w_i a_i + \frac{1}{\sigma_a^2} (\mathbf{A}^{-1} \mathbf{a})_i = w_i \left[ \frac{\delta_i}{w_i} - 1 + a_i \right]$$



Si ara definim el pseudo-registre com a  $y_i^* = \frac{\delta_i}{w_i} - 1 + a_i$ , obtenim la mateixa equació

48 que en el cas del model animal lineal:

$$w_i a_i + \frac{1}{\sigma_a^2} (\mathbf{A}^{-1} \mathbf{a})_i = w_i y_i^*$$

Per tant, veiem que els pseudo-registres de longevitat per l'animal  $i$ , sigui quina sigui la distribució base, són:

$$y_i^* = \frac{\delta_i}{w_i} - 1 + a_i \quad [49]$$

Els pesos associats  $w_i = H(y_i; \mathbf{x}_i; \mathbf{z}_i)$  són el risc acumulat de l'animal  $i$  del temps 0 fins al temps de *fallida* o censura.

### **Derivació dels pseudo-registres assumint una distribució de Weibull**

Quan la distribució base és una distribució de Weibull, els pseudo-registres de longevitat per l'animal  $i$  segueixen essent:

$$y_i^* = \frac{\delta_i}{w_i} - 1 + a_i \quad [50]$$

Els pesos associats  $w_i = H(y_i; \mathbf{x}_i; \mathbf{z}_i) = y_i^\rho \exp(\mathbf{x}_i' \boldsymbol{\beta}^* + \mathbf{z}_i' \mathbf{a})$  són el risc acumulat de la distribució de Weibull de l'animal  $i$  del temps 0 fins al temps de *fallida* o censura.

### **Història de l'Anàlisi de Supervivència en Producció Animal**

L'anàlisi de la longevitat i la supervivència ha rebut atenció creixent en diferents animals de producció, però on aquest subjecte ha adquirit més atenció ha estat en el boví lleter (Yazdi, 2000). Per tant, aquest bestiar ha estat pioner en el desenvolupament de l'aplicació de l'anàlisi de supervivència a la producció animal. El primer intent d'avaluar toros lleters a partir de mesures de la longevitat de les seves filles usant tècniques d'anàlisi de supervivència utilitzà un model de riscos proporcionals estratificat per l'any de naixement de la vaca, incloent com a únics efectes fixes el grup genètic i l'explotació (Smith, 1983). Posteriorment, la inclusió d'efectes aleatoris en el model de riscos proporcionals (Smith i Quaas, 1984) i,

particularment, l'extensió dels models de supervivència mixtes per incloure relacions entre progenitors (Ducrocq i Casella, 1996) i el desenvolupament de programes informàtics específics (Ducrocq i Sölkner, 1994, 1998) han fet possible estimar el potencial genètic dels toros per allargar la vida productiva de les vaques lleteres.

Considerant aquests desenvolupaments es va proposar una estratègia general per l'anàlisi de supervivència en millora genètica i epidemiologia (Ducrocq, 1997). Aquesta comença amb la realització d'una anàlisi preliminar de les dades basat en un model de riscos proporcionals semiparamètric i la comprovació de la hipòtesi de riscos proporcionals. Si la hipòtesi no és assumible, l'estratificació i/o l'ús de variables dependents de la vida productiva permetran obtenir un millor ajustament. Després, l'estimació no paramètrica de la funció de risc base ens proporciona un test gràfic per poder validar un model de Weibull (si aquest es valida, augmenta la precisió de les estimacions i es produeixen grans reduccions en el temps de computació). Quan s'hi inclouen efectes aleatoris, els paràmetres de variància són estimats a partir de l'aproximació de la densitat marginal posterior. Així, assumint que aquests components de variància són coneguts, un pot procedir a l'estimació dels efectes fixes i aleatoris (i als paràmetres de Weibull en el cas del model de Weibull) i a la computació de les funcions de supervivència esperades per cada progenitor. Això últim permet visualitzar la influència de cada variable sobre la longevitat, especialment quan són dependents de la vida productiva.

Seguint aquesta estratègia general, l'anàlisi de supervivència s'ha aplicat a l'estudi de la longevitat tant en boví de llet (Smith i Quaas, 1984; Ducrocq, 1987; Dürr i col·l., 1999; Vukasinovic, 1999; Vollema, 1998a), com en porcí (Brandt i col·l., 1999; Yazdi i col·l., 2000), en aus (Ducrocq i col·l., 2000), en cavalls (Wallin i col·l., 2000) i en oví (Southey i col·l., 2001). En boví de llet, es duen a terme avaluacions nacionals per al caràcter longevitat amb aquestes tècniques mitjançant l'ús del Survival Kit en un nombre creixent de països com França des del 1997 (Ducrocq, 1999a), Àustria i Alemanya des del 1998 (Ducrocq i Sölkner, 1998b; Pasman i Reinhardt, 1999a, 1999b), Holanda des del 1999 (De Jong i col·l., 1999) i Itàlia des del 1999 (Schneider i Miglior, 1999).

Fins al moment, els estudis sobre la longevitat i la supervivència dels vedells basats en l'esquema anterior en boví de carn són escassos i recents (Díaz i col·l., 2002; Rogers i col·l., 2004). Per tant, l'anàlisi de la longevitat de les vaques i la supervivència dels vedells en la *Bruna dels Pirineus* pot ser un exemple representatiu de les races de carn presents en les regions muntanyoses d'Espanya i d'Europa.

**Bibliografia**

- Aalen O.O. 1994. Effects of frailty in survival analysis. *Stat. Meth. Med. Res.* 3, 227–243.
- Achcar J.A., Bolfarine H. 1986. Use of accurate approximations for posterior densities in regression models with censored data. *Rev. Soc. Chil. Estad.* 3, 84-104.
- Allison P.D. 1995. *Survival Analysis Using the SAS® System. A Practical Guide*. SAS Institute Inc., Cary, North Carolina.
- Allore H.G., Warnick L.D., Hertl J., Gröhn Y.T. 2001. Censoring in survival analysis: a simulation study of the effect of milk yield on conception. *Prev. Vet. Med.* 49, 223-234.
- Anderson J.E., Louis T.A., Holm N.V., Harvald B. 1992. Time-dependent association measures for bivariate survival analysis. *J. Amer. Stat. Ass.* 87, 641-650.
- Berger J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, USA.
- Brandt H., Brevern N. von, Glodek P. 1999. Factors affecting survival rate of crossbred sows in weaner production. *Livest. Prod. Sci.* 57, 127-135.
- Breslow N.E. 1974. Covariance analysis of censored survival data. *Biometrics* 30, 89-99.
- Cantor A. 1999. *Extending SAS® Survival Analysis Techniques for Medical Research*. SAS Institute Inc. Cary, North Carolina.
- Clayton D.G. 1991. A Monte-Carlo method for Bayesian inference in frailty models. *Biometrics* 47, 467-485.
- Clayton D.G, Cuzick J. 1985. Multivariate generalizations of the proportional hazards model. *J. Royal. Stat. Soc. Series A* 148, 82-117.
- Cox D.R. 1972. Regression models and life tables (with discussion). *J. Royal Stat. Soc. Series B* 34, 187-220.
- De Jong G., Vollema A.R., Van der Beek S., Harbers A. 1999. Breeding value for functional longevity in the Netherlands. *Interbull Bulletin* 21, 68-72.
- Dellaportas P., Smith A.F.M. 1993. Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Appl. Stat.* 42, 443-459.
- Díaz C., Chirinos Z., Moreno A., Carabaño M.J. 2002. Preliminary analysis of functional longevity in the Avileña Negra Ibérica beef cattle breed. A: *Proceedings of 7<sup>th</sup> World Congress on Genetics Applied to Livestock Production*, Montpellier (França) 29, 697-700.

- Ducrocq V. 1987. *An analysis of length of productive life in dairy cattle*. Ph.D. thesis, Cornell University, Ithaca, New York, USA.
- Ducrocq V. 1997. Survival analysis, a statistical tool for length of productive life data. A: *Proceedings of the 48<sup>th</sup> EAAP Meeting*, Vienna (Austria).
- Ducrocq V. 2000. *Survival analysis applied to animal breeding and epidemiology*. Course notes. 86 pp.
- Ducrocq V. 2001 A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life. *Interbull Bulletin* 27, 147-152.
- Ducrocq V. 2004. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. A: *Proceedings of the 55th EAAP Meeting*, Bled (Slovenia), Paper G6.11.
- Ducrocq V. 1999a. Extension of survival analysis models to discrete measures of longevity. *Interbull Bulletin*, 20: 41-47.
- Ducrocq V. 1999b. Two years of experience with the French genetic evaluation of dairy bulls on production-adjusted longevity of their daughters. *Interbull Bulletin*, 20: 60-67.
- Ducrocq V., Casella G. 1996. A Bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 28, 505-529.
- Ducrocq V., Sölkner J. 1994. "The Survival Kit", a FORTRAN package for the analysis of survival data. A: *Proceedings of the 5<sup>th</sup> World Congress on Genetics Applied to Livestock Production*, Univ. of Guelph, Ontario (Canada), 22, 51-52.
- Ducrocq V., Sölkner J. 1998a. "The Survival Kit v3.12", a FORTRAN package for large analysis of survival data, A: *Proceedings of the 6<sup>th</sup> World Congress on Genetics Applied to Livestock Production*, University of New-England, Armidale (Australia), 23, 447-450.
- Ducrocq V., Sölkner J. 1998b. Implementation of a routine breeding value evaluation for longevity of dairy cows using survival analysis techniques. A: *Proceedings of the 6<sup>th</sup> World Congress on Genetics Applied to Livestock Production*, University of New-England, Armidale (Australia), 23, 359-362.
- Ducrocq V., Quaas R.L., Pollak E.J., Casella G. 1988a. Length of productive life of dairy cows. 1. Justification of a Weibull model. *J. Dairy Sci.* 71, 3061-3070.
- Ducrocq V., Quaas R.L., Pollak E.J., Casella G. 1988b. Length of productive life of dairy cows. 2. Variance component estimation and sire evaluation. *J. Dairy Sci.* 71, 3070-3079.

- Ducrocq V., Besbes B., Protais M. 2000. Genetic improvement of laying hens viability using survival analysis. *Genet. Sel. Evol.* 32, 23-40.
- Ducrocq V., Boichard D., Barbat A., Larroque H. 2001 Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index. A: *Proceedings of the 52nd EAAP Meeting, Budapest (Hongria)*, Paper G1.4.
- Dürr J.W., Monardes H.G., Cue R.I. 1999. Genetic Analysis of herd life in Quebec Holstein using Weibull models. *J. Dairy Sci.* 82, 2503-2513.
- Goutis C., Casella G. 1996. *Explaining the saddlepoint approximation*. Technical report, Biometrics Unit, Cornell University, Ithaca, NY, USA, BU-1311-M.
- Hougaard P. 1986a. Survival models for heterogeneous populations derived from stable distributions. *Biometrika* 73, 387-396.
- Hougaard P. 1986b. A class of multivariate failure time distributions. *Biometrika* 73, 671-678.
- Kalbfleisch J.D., Prentice R.L. 1980. *The statistical analysis of failure time data*. Wiley, New York.
- 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.
- Klein J.P., Moeschberger M., Li Y.H., Wang S.T. 1992. Estimation random effects in the Framingham heart study. A: *Survival Analysis: State of the Art*. J. Klein i P. Goel (ed), Kluwer Academic Publishers, pp. 99-120.
- Klein J.P., Moeschberger M. 1997. *Survival analysis*. Springer-Verlag, New York, USA.
- Kleinbaum D.G. 1996. *Survival analysis: a self-learning text*. Springer, New York, USA.
- Korsgaard I.R. 1996. Genetic analysis of survival data-a Gibbs sampling approach. *Interbull Bulletin* 12, 161-169.
- Korsgaard I.R., Madsen P., Jensen J. 1998. Bayesian inference in the semiparametric log normal frailty model using Gibbs sampling. *Genet. Sel. Evol.* 30, 241-256.
- Korsgaard I.R., Andersen A.H., Jensen J. 1999. Discussion of heritability of survival traits. *Interbull Bulletin* 20, 31-35.
- Korsgaard I.R., Andersen A.H., Jensen J. 2000. On different models, on heritability, reliability and related quantities of survival traits. A: *Proceedings of the 51st EAAP Meeting, Wageningen (Holanda)*.

- Korsgaard I.R., Andersen A.H., Jensen J. 2002. Prediction error variance and expected response to selection, when selection is based on the best predictor for Gaussian and threshold characters, traits following a Poisson mixed model and survival traits. *Genet. Sel. Evol.* 34, 307-333.
- Louis T. 1991. Assessing, accomodating and interpreting the influences of heterogeneity. *Environ. Health Perspect.* 80, 215-222.
- Pasman E., Reinhardt F. 1999a. Genetic evaluation for length of productive life of Holstein cattle in Germany. *Interbull Bulletin* 21, 55-59.
- Pasman E., Reinhardt F. 1999b. Genetic relationships between type composites and length of productive life of Black-and-White Holstein cattle in Germany. *Interbull Bulletin* 21, 117-121.
- Peto R. 1972. Contribution to the discussion of paper by D.R. Cox. *J. Royal Stat. Soc. Series B* 34,205-207.
- Robert C. 1992. *L'Analyse Statistique Bayésienne*. Economica, Paris, France.
- Rogers P.L., Gaskins C.T., Johnson K.A., MacNeil M.D. 2004. Evaluating longevity of composite beef females using survival analysis techniques. *J. Anim. Sci.* 82, 860-866.
- Schneider M.P., Miglior F. 1999. A proposal for genetic evaluation for functional herd life in Italian Holsteins. A: *Proceedings of the 50th EAAP Meeting*, Zurich (Switzerland), 5, 11-15.
- Smith S.P. 1983. *The extension of failure time analysis to problems of animal breeding*. Ph.D. thesis, Cornell University, Ithaca, New York, USA.
- Smith S.P., Quaas R.L. 1984. Productive life span of bull progeny groups: failure time analysis. *J. Dairy Sci.* 67, 2999-3007.
- Smith A., Skene A.M., Shaw J., Naylor J., Dransfield M. 1985. The implementation of the Bayesian paradigm. *Communic. Statist. – Theor. Meth.* 14, 1079-1102.
- Southey B.R., Rodriguez-Zas S.L., Leymaster A.K. 2001. Survival analysis of lamb mortality in a terminal sire composite population. *J. Anim. Sci.* 79, 2298-2306.
- Tempelman R.J., Gianola D. 1993. Marginal maximum likelihood estimation of variance components in Poisson mixed models using Laplacian integration. *Genet. Sel. Evol.* 25, 305-319.
- Tempelman R.J., Gianola D. 1994. Estimation of genetic variance for overdispersed count data with an animal breeding application. A: *Joint Statistical Meetings*, Toronto (Canada).

- Tierney L., Kardane J.B. 1986. Accurate approximations for posterior moments and marginal densities. *J. Amer. Stat. Ass.* 81, 82-86.
- Tierney L., Kass R.E., Kardane J.B. 1989. Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Stat. Ass.* 84, 710-716.
- Tsiatis A.A. 1981. A large sample study of Cox's regression model. *Ann. Stat.* 9, 93-108.
- Vaupel J., Manton K.G., Stallard E. 1979. The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography* 16, 439-454.
- Vollema A.R. 1998a. *Selection for longevity in dairy cattle*. PhD. Thesis, Wageningen Agricultural University, Wageningen (Holanda).
- Vukasinovic N. 1999. Application of survival analysis in breeding for longevity. *Interbull Bulletin* 21, 3-15.
- Wallin L., Strandberg E., Philipson J., Dalin G. 2000. Estimates of longevity and causes of culling and death in Swedish warmblood and coldblood horses. *Livest. Prod. Sci.* 63, 275-289.
- Yazdi M.H., Rydhmer L., Ringmar-Cederberg E., Lundeheim N., Johansson K. 2000. Genetic study of length of productive life in Swedish Landrace sows. *Livest. Prod. Sci.* 63, 255-264.
- Yazdi M.H., Visscher P.M., Ducrocq V., Thompson R. 2002. Heritability, reliability of genetic evaluations and response to selection in proportional hazards models. *J. Dairy Sci.* 85, 1563-1577.





## **Capítol 2**

### **FACTORS INFLUENCING LENGTH OF PRODUCTIVE LIFE AND REPLACEMENT RATES IN THE BRUNA DELS PIRINEUS BEEF BREED**

**J. Tarrés<sup>a</sup>, P. Puig<sup>b</sup>, V. Ducrocq<sup>c</sup>, J. Piedrafita<sup>a</sup>**

<sup>a</sup> *Grup de Recerca en Remugants, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*

<sup>b</sup> *Departament de Matemàtiques, Servei d'Estadística, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*

<sup>c</sup> *Station de Génétique Quantitative et Appliquée, Institut National de la Recherche Agronomique, 78352 Jouy-en-Josas, France*

***Animal Science 78 (1), 13-22 (2004)***

## **Abstract**

An analysis of the length of productive life in Bruna dels Pirineus beef breed was performed with a non-parametric approach giving an average value of 9 years of productive life, and a corresponding replacement rate of 11%. Using a proportional hazards model stratified by herd, the influence of calf birth weight and weight gain until weaning, calving difficulty, calving interval and age at first calving on length of productive life was studied. Two models were explored: the first one included time-dependent variables taking the current value of the covariate at each calving date, while the second one also comprised time-dependent interactions between the value of the covariate of the current calving and its mean value during last (up to) three previous calvings. Results from the first model showed that the risk of culling increases with very high ages at first calving, increasing calving difficulties, very large calf birth weights, very small weight gains until weaning and very long calving intervals. Furthermore, results from the second model showed that these increases also depend upon the sequence of values for the same covariate in previous calvings. Finally, these higher risks of culling implied lower survival functions that increased replacement rates but only slightly decreased average performances.

**Keywords:** beef cattle, productive life, survival analysis, replacement rate.

## **Introduction**

Culling is one of the most complex decisions a farmer has to take to make the herd more profitable. A long herd life of a dairy cow substantially decreases replacement costs and enables her to achieve her maximum performance capacity when attaining full maturity (Essl, 1998), so several models have been developed to maximise the net present value from current and replacement cows over a given decision horizon (Rajala-Schultz and Gröhn, 2001). Furthermore, results of several studies have shown that long lifetime production and low culling rates in swine herds have substantial economic benefits. Thus, length of productive life is a trait which has received increasing attention in animal breeding (Yazdi et al., 2000). However, little is known on the length of productive life of beef cattle and its consequences on the productivity of the herd.

The lifespan of a female farm animal can be partitioned into two time periods: a) a costly raising period from birth to first parity, and b) the subsequent reproductive period until disposal (Essl, 1998). This second period is generally called length of productive life, and the capability of an animal to remain in its herd over time is often called stayability or survival. Ducrocq et al. (1988) have proposed to distinguish between “true” and “functional” longevity. True longevity is defined as the ability to delay culling, whereas functional longevity would be the ability to delay involuntary disposal. Nevertheless, there is no direct measure of functional longevity, because voluntary culling always exists to some extent in livestock populations and in some species (e.g., swine and dairy cattle) it is often difficult to distinguish between true and functional length of productive life due to insufficient recording systems for this purpose.

Ducrocq (1997) proposed a general strategy for the analysis of length of productive life in animal breeding and epidemiology based on survival analysis as an adequate method for the genetic evaluation of length of productive life measurements. This methodology has been applied in several farm species, although only scarcely in beef breeding so far (Tarrés et al., 2002; Díaz et al., 2002). In this paper several factors influencing length of productive life in the Bruna dels Pirineus beef breed, that could be a representative example for the beef cattle breeds present in the mountainous areas of Europe, were analysed. These factors were age at first calving, calf birth weight, calving difficulty, calf weight gain until weaning and calving interval. After estimating the observed survival function, its dependence on culling factors was investigated using a

Proportional Hazards Model. Furthermore, expected survival functions were calculated in different hypothetical situations accounting for this dependence, as well as replacement rates and average performances.

## **Material and methods**

### ***Data***

Bruna dels Pirineus is a beef breed selected from the old Swiss Brown (derived from Schwyz canton), similar to the current American Braunvieh, and located in the eastern part of the Pyrenees. Data were recorded between the years 1983 and 2002 in three breeding herds participating in the Yield Recording Scheme of the breed, and included data from 1331 beef cows having 5790 calves. The pedigree of the cows was only partly known, as 53 bulls sired 834 cows, the rest of the sires being unknown. The length of the productive life of a cow was estimated as the difference between the date of culling and the date at first calving, divided by 365.25 days. When the date of culling was unknown, we assumed that the cow was censored 180 days after the date of the last birth registered. These 180 days were the period of time after calving in which cows usually are kept in the herd to rear their calves. After editing, our database included data of the productive life of 1061 cows, with 436 complete records (41.1%) and 625 censored records (58.9%).

### ***Survival analysis***

The distribution of lengths of productive life was described by the Kaplan-Meier estimate of the survival function (Kaplan and Meier, 1958). This observed distribution was the result of the influence of different factors involved in the culling process. To investigate the influence of these factors on risk of culling, a semiparametric proportional hazards model was assumed (Cox, 1972). The analysis was carried out using the package 'The Survival Kit v3.12' (Ducrocq and Solkner, 1998). The culling factors analysed were age at first calving, calf birth weight, calf weight gain until weaning, calving difficulty and calving interval. Calf weight gain until weaning was chosen instead of weaning weight to avoid a high correlation between covariates. Factors with repeated data were included in the analysis as time dependent variables.

The importance of each factor was assessed by a sequential likelihood ratio test (Ducrocq and Solkner, 1998).

The first operational model assumed was the “current calving value model” (Model 1):

$$h_g(t; X) = h_{0g}(t) \exp[AFC_i + P * CD_j + P * BW_k + P * WG_l + P * CI_m + HY] \quad [1]$$

where age at first calving (AFC) was treated as a time independent categorical variable and calving difficulty (CD), calf birth weight (BW), calf weight gain until weaning (WG) and calving interval (CI) were treated as time dependent categorical variables changing at every calving date. The effects for primiparous and multiparous cows were assumed to be different and were modelled by including an interaction with a parity indicator (P) taking two values. It initially takes a value 1 and changes to value 2 at the second calving and after. This model also allowed a different baseline hazard functions for each herd  $h_{0g}(t)$  (g: 1 to 3) allowing to relax the assumption of proportional hazards across herds. Indeed, non-proportionality of the hazards can be suspected, since culling decisions usually are different among herds along productive life. Furthermore, it is known that the risk of culling can change every year due to several sources of variation such as natural forage availability, changes in herd management practices and other environmental circumstances. The inclusion of an interaction term between herd and year allows a precise modelling of the changes in the culling policies on a farm over time (Ducrocq, 1997). The herd×year effect (HY) was treated as a time dependent fixed variable changing on January first, each year.

All variables in the model were included as categorical variables. CD had six classes: calving without assistance (1); calving slightly assisted by the farmer (2); calving strongly assisted by the farmer (3); calving assisted by the veterinary practitioner (4); caesaria (5); and missing values (miss). BW, WG, CI and AFC, originally which are continuous variables, were transformed into categorical variables within six arbitrary classes: very small (vs – bottom 10% of the data), small (s- 23.3%), medium (m- 33.3%), large (l- 23.3%), very large (vl- top 10%), and missing values (miss). Cutpoints for the time dependent variables varied with the age of the cow at calving defined in terms of productive life (Table 1).

**Table 1.** *Cutpoints for different productive life intervals for the categorisation of independent variables characterising the current calving*

	Percentiles	Productive life interval (days)				
		0 – 300	300 – 600	600 - 1800	1800 - 3600	> 3600
Birth	10 %	36.6	39.1	40.1	41.1	40.1
Weight (kg)	33 %	41.1	42.7	44.9	45.1	43.9
	67 %	46.7	47.7	48.9	48.7	47.7
	90 %	51.9	52.9	53.9	53.9	52.9
Weight	10 %	155.1	155.1	160.1	166.1	166.1
Gain until	33 %	173.1	172.1	178.1	183.1	179.9
Weaning (kg)	67 %	190.3	189.5	197.9	201.9	197.1
	90 %	207.9	213.9	217.9	219.9	216.9
Calving Interval (days)	10 %		346	333	332	336
	33 %		370	354	354	358
	67 %		404	382	386	391
	90 %		453	438	449	450

Model 1 assumed that the risk of culling depended on the current value of a factor in each calving. However, the values of the variables of previous calvings might also influence the risk of culling in a certain calving (Tarrés et al., 2002). So a second operational model including a simple summary of the information on previous calvings was fitted. The interaction model between current calving value and mean of previous calvings (Model 2) was:

$$h_g(t; X) = h_{0g}(t) \exp \left[ AFC_i + P * CD_j * pCD_n + P * BW_k * mBW_o + \right. \quad [2] \\ \left. + P * WG_l * mWG_p + P * CI_m * mCI_q + HY \right]$$

where age at first calving (AFC) was treated as a time independent categorical variable. As in Model 1, this model allowed different baseline hazard functions on each herd  $h_{0g}(t)$  (g: 1 to 3) and herd×year effect (HY) was treated again as a time dependent fixed variable changing on January first each year. The effect of calving difficulty was included as the interaction term between parity (P), calving difficulty score in the current calving (CD) and calving difficulty in the previous calving (pCD) treated as a time dependent categorical variable changing at every calving date. Previous calving difficulty (pCD) is obviously ignored for all first calving, and takes the value of the calving difficulty of first calving for the second calving, the calving difficulty of second

calving for the third calving and so on. However, to avoid having a large number of categories with few records each, pCD was grouped into three classes: calving without assistance and calving slightly assisted by the farmer (“easy”); calving strongly assisted by the farmer, calving assisted by the veterinary practitioner and caesaria (“difficult”); and missing values (“miss”). The effects of calf birth weight (BW), calf weight gain until weaning (WG) and calving interval (CI) were included as the interaction term between the parity (P), the value in the current calving (BW, WG and CI) and the mean value of the trait in the (up to) three previous calvings (mBW, mWG and mCI – when available) treated as time dependent categorical variables changing at every calving date. For example, mBW is ignored for all first calvings, and includes the first calf birth weight for the second calving, the mean of two first calf birth weights for the third calving, the mean of the three first calf birth weights for the fourth calving, the mean of second, third and fourth calf birth weights for the fifth calving and so on. The categorisation of continuous previous means was similar to that applied in Model 1, and the cutpoints for the means are shown in Table 2. After the categorization, adjacent classes with less than 50 elementary records were grouped together.

**Table 2.** *Cutpoints for different productive life intervals for the categorisation of independent mean variables characterising previous calvings*

		Productive life interval (days)				
	Percentiles	0 – 300	300 – 600	600 - 1800	1800 - 3600	> 3600
Birth	10 %	36.6	38.1	41.1	42.1	39.6
Weight (kg)	33 %	41.1	42.4	45.4	46.4	43.9
	67 %	46.7	46.4	49.4	50.4	47.9
	90 %	51.9	50.9	53.9	54.9	52.4
Weight	10 %	155.1	155.6	161.6	166.6	163.6
Gain until	33 %	173.1	172.6	178.6	183.6	180.6
Weaning (kg)	67 %	190.3	190.9	196.9	201.9	198.9
	90 %	207.9	207.6	213.6	218.6	215.6
Calving Interval (days)	10 %		346	354.4	347.4	352
	33 %		370	368.9	361.9	366.9
	67 %		404	386.1	379.1	384.1
	90 %		453	406	399	404

### **Expected survival functions**

Once the regression coefficients  $\hat{\beta}$  for the culling factors had been estimated maximising Cox' partial likelihood, relative hazard ratios (HR) of cows having a particular set of characteristics can be estimated as  $HR = \exp(\mathbf{X}\hat{\beta})$ . The baseline survival functions  $\hat{S}_{0,g}(t)$  are computed using an empirical approach similar to the one leading to the Kaplan-Meier estimate. These empirical baseline survival functions allowed us to obtain expected survival functions in an hypothetical herd with all cows having a particular sequence of covariates as  $\hat{S}(t,g,X) = [\hat{S}_{0,g}(t)]^{HR}$ .

### **Replacement rates**

The replacement rate associated with any expected survival function was computed with the following steps:

(i) Consider a herd with  $N$  animals distributed in  $k+1$  distinct and ordered intervals of productive life,  $(t_i, t_{i+1})$ :  $N = n_0 + n_1 + n_2 + \dots + n_k = \sum_{i=0}^k n_i$ , with  $n_i$  being the number of animals included in the interval  $(t_i, t_{i+1})$ . The upper bound  $t_{k+1}$  of the  $k^{th}$  productive life interval was the longest desired utilisation time of cows. In our study, this was assumed to be equal to the maximum observed longevity, i.e. 16 years.

(ii) This figure  $n_i$  can be computed as  $n_i = n \int_{t_i}^{t_{i+1}} S(t) dt \approx n S(t_i) \Delta t_i$ , where  $n$  is the number of replacement animals entering the herd each year,  $S(t_i)$  is the survival probability at  $t_i$  and  $\Delta t_i = t_{i+1} - t_i$  is the length of the interval of the productive life in years  $(t_i, t_{i+1})$ . So,  $N = n \sum_{i=0}^k S(t_i) \Delta t_i$ .

(iii) If the replacement rate (RR) is defined as the ratio between the number of animals entering the herd  $n$  in a particular year and the number of animals in the herd  $N$  and assuming constant herd size, then:

$$RR(k) = \frac{n}{N} = \frac{1}{\sum_{i=0}^k (S(t_i) \Delta t_i)} \quad [3]$$



This formula was used to compute the RRs from an empirical survival function.

(iv) When the length of the intervals of productive life are short enough, then the RR is inversely proportional to the expectation of the length of productive life  $E(t)$ :

$$RR(k) = \frac{n}{N} = \frac{1}{\int_0^{t_{k+1}} S(t) dt} = \frac{1}{E(t)}, \quad [4]$$

because  $E(t) = \int_0^{t_{k+1}} S(t) dt$  (Kalbfleish and Prentice, 1980).

**Prediction of average weaning weights due to the proportion of cows in each productive life interval**

One of the potential economical benefits of a longer productive life is the relative reduction of the importance of the first parity: a higher survival function implies a higher percentage of cows achieving their maximum capacity of performance when attaining full maturity. This aspect can be quantified by computing the average weaning weight of a herd weighted by the proportion of cows in each productive life interval (AWW). This was done in the following way:

(i) Calf weaning weights  $WW(t_i)$  depend on the parity of the cow. Then, the average weaning weight of a hypothetical herd depends on the proportion of cows in the different productive life intervals. So, the AWW is:

$$AWW = \frac{\sum_{i=0}^k n_i WW(t_i)}{N}$$

with  $n_i$  being the number of cows included in the interval  $(t_i, t_{i+1})$ . The  $WW(t_i)$  values used were those reported in Bruna dels Pirineus by Casellas and Piedrafita (2002).

(ii) The value of  $n_i$  depends on the survival function  $n_i \approx n S(t_i) \Delta t_i$ . So, the average weaning weight (AWW) is:

$$AWW = \frac{n \sum_{i=0}^k S(t_i) \Delta t_i WW(t_i)}{N} = RR(k) \left[ \sum_{i=0}^k S(t_i) \Delta t_i WW(t_i) \right] \quad [5]$$

## **Results and discussion**

### ***Observed length of productive life***

The observed survival function in each herd estimated by the Kaplan-Meier method is presented in Figure 1. Herd 1 and herd 2 have similar survival functions. The rate of decrease in the observed probability of survival was fairly constant until 10 years of productive life, but later it was slightly more pronounced. The estimated average length of productive life is respectively 8.9 and 9.0 years and the corresponding annual replacement rate was 11.3 % and 11.1%. In herd 3, culling rate is higher than in other herds (probably related to a harder environment) and so the probability of survival is lower. This last herd only has records from 1990 to 2000 and thus the maximum observed longevity is 10 years. For this reason, the replacement rate in herd 3 was not calculated because it would be overestimated. The full survival function in herd 3 could be estimated using truncated data, for example, data of animals with first calvings before 1990 but still productive in 1990 and after. However, this method requires to know the date of first calving of the cows, which was unknown unfortunately in our study.

**Figure 1.** *Observed within herd survival functions using the Kaplan-Meier formula*



Estimates of average length of productive life and RR in herds 1 and 2 are close to other estimates reported in beef cattle populations (Buxadé, 1997 and Díaz et al., 2002) and other species of extensive production, like meat sheep and goats (Torrent, 1991). Average length of productive life in beef cattle is substantially greater than that of other species reared more intensively, such as dairy cattle (nearly three years of productive life; Veerkamp et al., 1999) and swine (around 2.5 – 2.8 years; Yazdi et al., 2000).

### **Factors influencing length of productive life**

Age at first calving, calving difficulty, calf birth weight (NS for model 1), calf weight gain until weaning and calving interval were significantly related to length of productive life in both models: current calving value model (Model 1) and the model including the interaction between current calving value and mean of previous calvings (Model 2), as indicated by the likelihood estimates (Table 3). The herd-year effect had a strong influence on the culling process as indicates the sequential likelihood values (196.1 and 194.9;  $p < 0.001$ ) corresponding to the two models (Table 3). No estimates of the corresponding hazard ratios are shown here because they did not follow any consistent trend.

**Table 3.** Estimates of the sequential likelihood ratio tests for each covariate

Model 1			Model 2		
Variable	DF	Sequential LRT	Variable	DF	Sequential LRT
AFC	5	20.9 (***)	AFC	5	20.9 (***)
CD	11	193.8 (***)	CD*pCD	17	202.2 (***)
BW	11	11.2 (NS)	BW*mBW	35	50.3 (*)
WG	11	53.7 (***)	WG*mWG	35	89.5 (***)
CI	12	25.9 (*)	CI*mCI	36	70.3 (***)
HY	34	196.1 (***)	HY	34	194.9 (***)
<b>Model 1</b>	<b>84</b>	<b>501.7</b>	<b>Model 2</b>	<b>162</b>	<b>628.0</b>

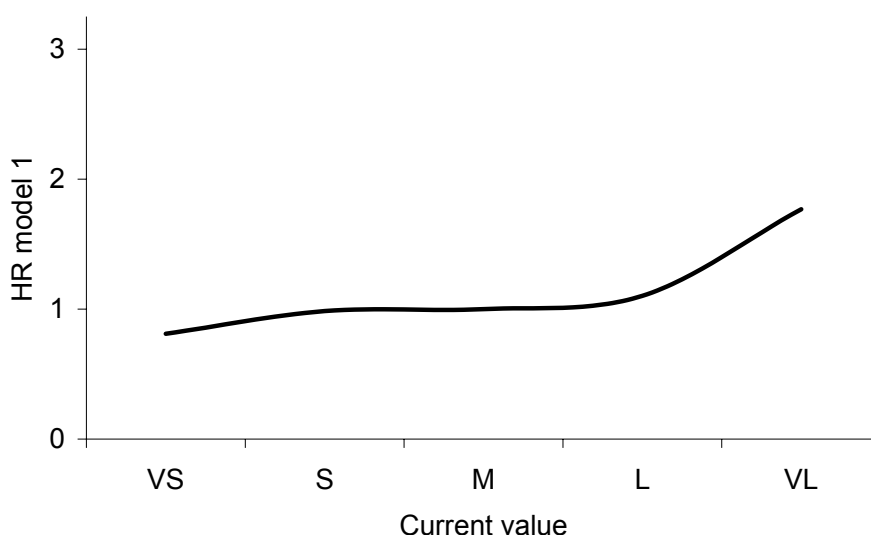
(AFC: Age at first calving; CD: Calving difficulty; BW: Calf birth weight; WG: Calf weight gain until weaning; CI: Calving interval; HY: Herd-Year)

(\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ ; NS non-significant)

Age at first calving had a significant effect on the culling process in both models ( $p < 0.05$ ). Cows with an age at first calving between 30 and 33 months had the

minimum risk of being culled, which increased as the cows get older at first calving (Figure 2). These results are comparable to those observed in other beef cattle breed (Díaz et al., 2002). In dairy cows, an increasing hazard ratio was also found in cows having their first calving older than 30 months (Essl, 1998). This would suggest that the first calving should not be postponed past 33 months because the higher risk of being culled will lead to higher replacement costs. It is important to note that the range for age at first calving analysed in this study was limited (30 to 42 months) and did not allow to assess the effects on the length of productive life of early first calvings (24 months) practised in some other herds of the breed.

**Figure 2.** Effect of age at first calving (AFC) on culling expressed as relative hazard ratio (HR) (reference = M)

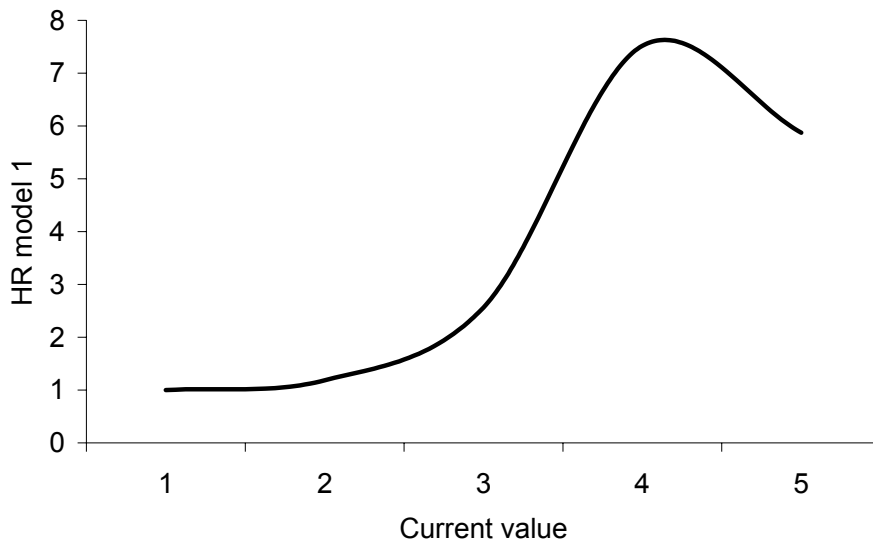


(VS = Very small, S= Small, M = Medium, L = Large, VL = Very large)

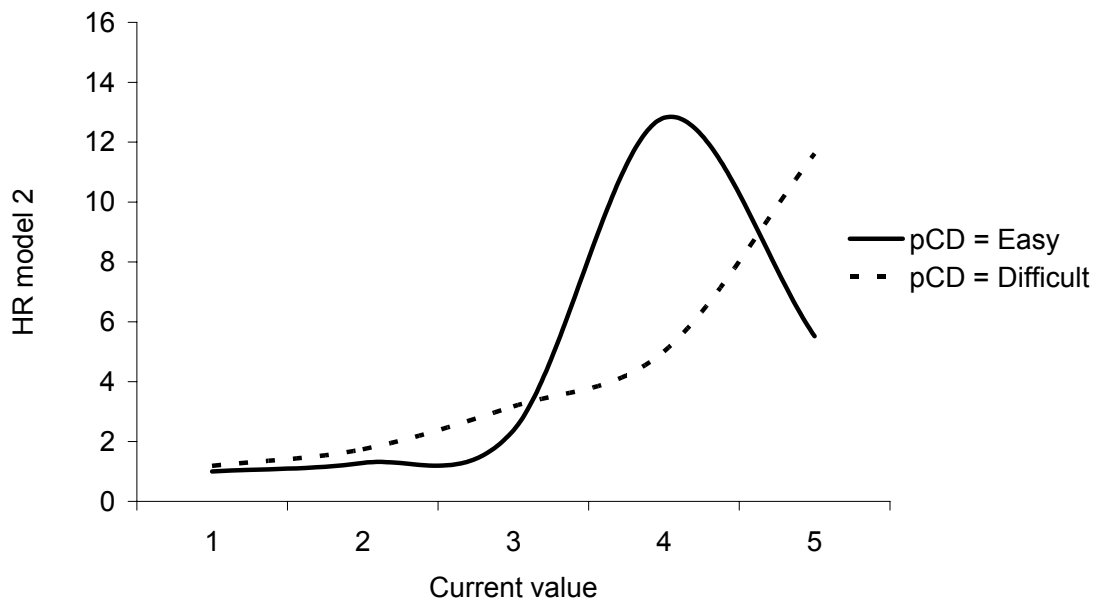
Calving difficulty also had a large influence on length of productive life (Table 3). Results from Model 1 showed that cows with non-assisted calving had minimum hazard values and the hazard increased as calvings became more difficult (Figure 3a). When the farmer’s assistance was very strong, cows were at a risk of being culled twice the one of cows with non-assisted calvings ( $p < 0.001$ ). When the intervention of the veterinary practitioner was needed, with or without caesaria, the hazard became six to eight times higher ( $p < 0.001$ ). Luo et al. (2002) also reported that calving difficulty

**Figure 3.** Effect of calving difficulty (CD) on culling expressed as relative hazard ratio (HR) (reference=1)

(a) Model 1



(b) Model 2, including an interaction with the previous calving (pCD)



(1 = no assistance, 2 = slightly assisted, 3 = strongly assisted, 4 = assisted by the veterinary practitioner, 5 = caesaria)

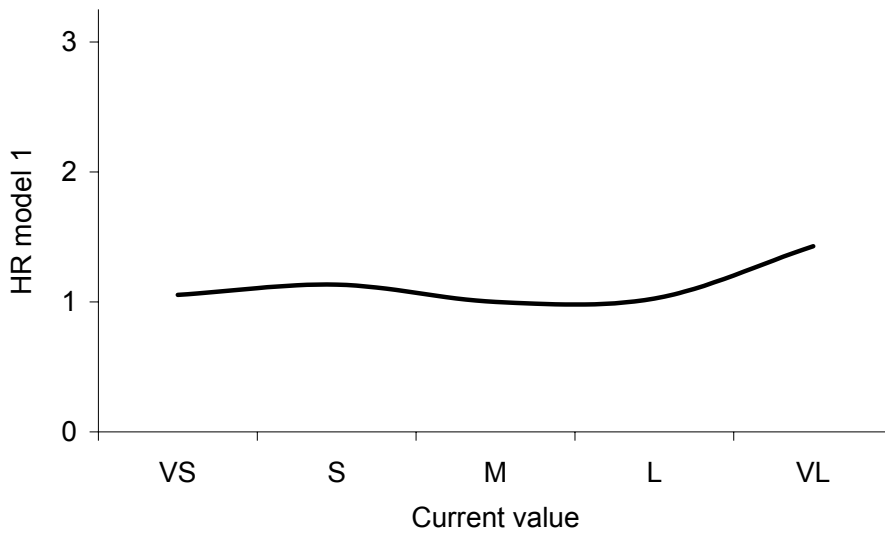
as well as a large birth weight, augmented culling rate in Canadian Holsteins. As a consequence, any decision to improve calving ease could reduce replacement costs, as well as direct losses owing to dystocia, stillbirth, veterinary costs and labour requirements. On the other hand, there were different HR within the same class of CD depending on pCD (Figure 3b), which would suggest that previous calving difficulty might influence the risk of culling by interacting with current calving difficulty. However, the interactions did not reach statistical significance in our data set, probably because there were too many classes with few records. Further research would be needed to assess the relevance of this putative interaction effect.

Calf birth weight had a non-significant influence in the culling process in model 1 (Table 3), maybe because the effect of calving difficulty was accounted for in the model. Cows with intermediate calf birth weights had the lowest risk of being culled. They were used to define the reference value in Figure 4a. Results from Model 1 showed that this baseline hazard slightly increased when calf birth weights were very large ( $p < 0.10$ ). However, results from Model 2 showed that the increase in the HR also depended upon the mean of previous calvings (Figure 4b). A cow having a calf with a very large birth weight had twice more risk of being culled if the previous mean was small or medium ( $p < 0.05$ ), but it had no effect if previous mean was also very large. Furthermore, when the mean of previous BWs was very large, the risk of being culled decreased if the current birth weight was not very large ( $p < 0.05$ ). These results have been consistent irrespective of the inclusion or not in the model of any of the other culling factors. One possible explanation is that there could have been some involuntary culling due to heavy weight of calf at birth in each calving, but the cow might become adapted to elevated birth weights depending upon the previous sequence of weights at birth. Despite it could be other possible explanation to these effects, what is noticeable is that they were robust to the exclusion or the inclusion in the model of anyone of the other culling factors.

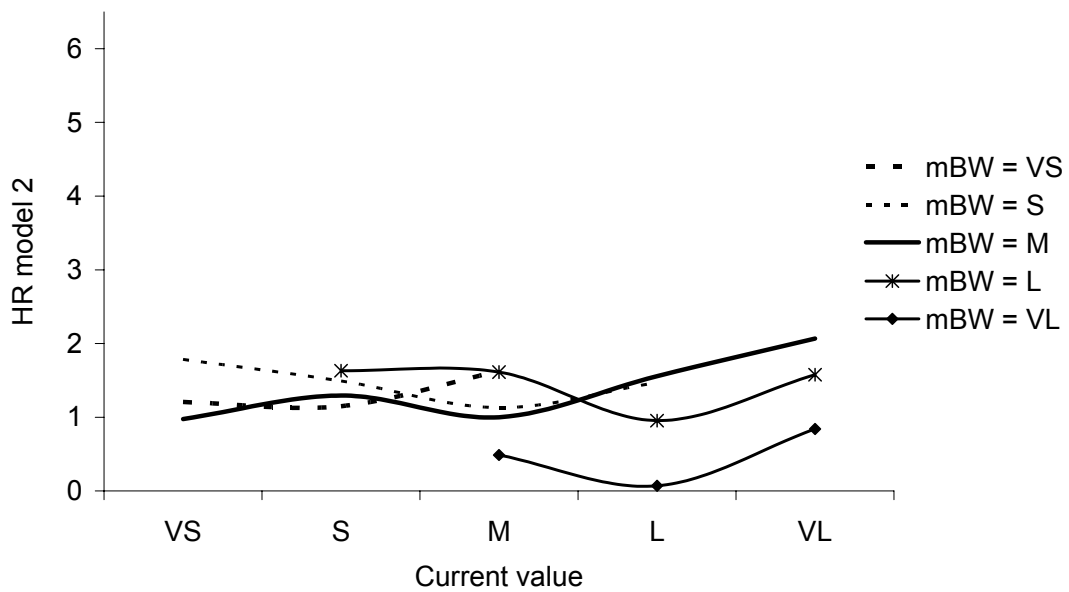
Weight gain until weaning also influenced longevity (Table 3). Cows which produced calves with intermediate weight gain until weaning had the lowest risk of being culled, and were chosen as reference in Figure 5a. Results from Model 1 showed that the hazard slightly increased when weight gain until weaning was very small ( $p < 0.10$ ). However, results from Model 2 showed that the increase in hazard also depended on mWG (Figure 5b). Very small WG had no effect on culling if mWG was intermediate or large but it had an important effect if mWG was small or very small ( $p < 0.05$ ). Voluntary culling of cows having calves with very small weight gains was

**Figure 4.** Effect of calf birth weight (BW) on culling expressed as relative hazard ratio (HR) (reference = M)

(a) Model 1



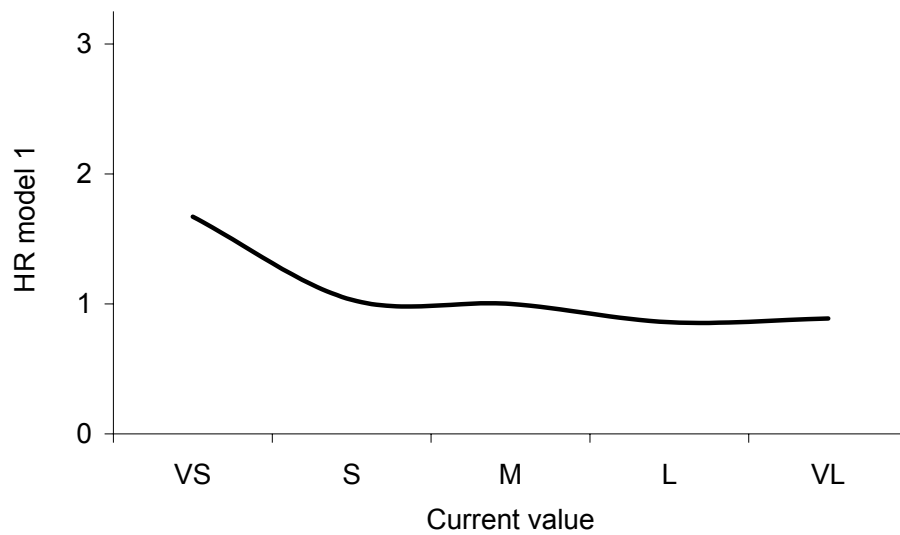
(b) Model 2, including an interaction with the mean of previous calvings (mBW)



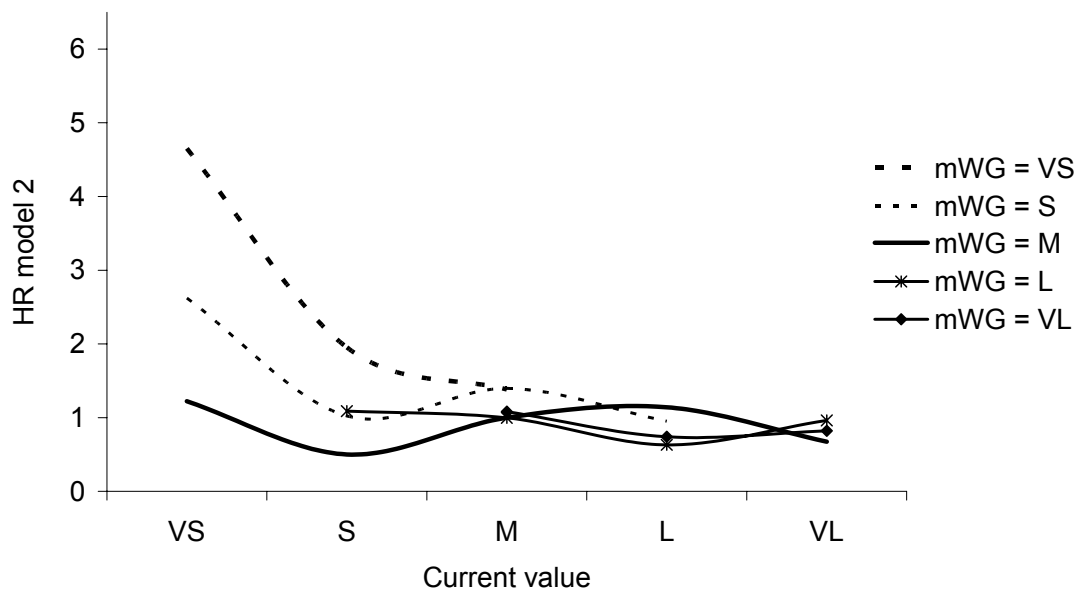
(VS = Very small, S= Small, M = Medium, L = Large, VL = Very large)

**Figure 5.** Effect of weight gain until weaning (WG) on culling expressed as relative hazard ratio (HR) (reference = M)

(a) Model 1



(b) Model 2, including an interaction with the mean of previous calvings (mWG)



(VS = Very small, S= Small, M = Medium, L = Large, VL = Very large)



clearer than those producing calves with small birth weight, probably because weight gain has a greater impact on the saleable product, the weaned calf. Similar results are observed in other beef cattle breeds where a low level of calf weaning weight is a reason for culling cows (Díaz et al., 2002). Voluntary culling caused by poor production has also been described for milk production in dairy cattle (Ducrocq et al., 1988; Rajala-Schultz and Gröhn, 2001). In swine, a higher risk of culling in sows with small litter size in the first and last parities has been reported (Yazdi et al., 2000). On the contrary, and similarly to what is described by these authors, Model 1 and Model 2 did not detect an increased risk of culling caused by higher calf performances, which suggests that the Bruna dels Pirineus breed may have a good maternal ability which permits the cow to rear a calf with little effort.

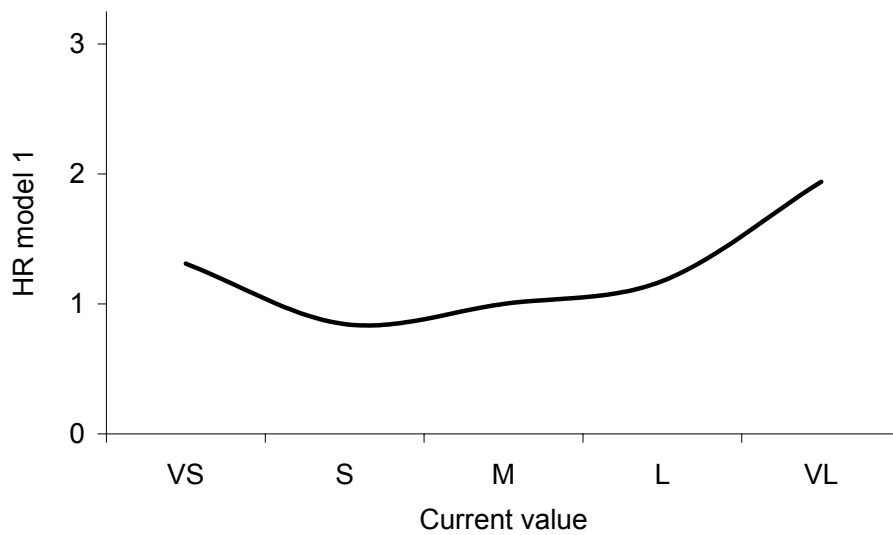
Again, cows with intermediate calving intervals had the lowest risk of being culled and were used as reference in Figure 6a. Results from Model 1 showed that only very long calving intervals, probably caused by reproductive problems, increased hazard significantly ( $p < 0.001$ ). Furthermore, results from Model 2 showed that the increase in the HR depended significantly upon the interaction with the mean of previous calvings. The hazard of very long CI was especially higher when the mCI was also very long (Figure 6b). Similarly to what happens in species of more intensive production systems like sows (Yadzy et al., 2000), the results showed that the increase of hazard due to reproductive problems in beef cattle in extensive production systems could be important. Moreover, results from Model 2 suggest that the accumulation of calving intervals shorter than a year, as it occurs by intensifying production, could double the risk of being culled when the current calving interval was shorter than 11 months (Figure 6b). However, this increase of hazard only approached statistical significance ( $p < 0.10$ ), and further research would be needed to assess the relevance of this effect.

Finally, the missing values corresponded to a much higher risk of being culled in both models. This can be attributed to the fact that often the cow is at a higher risk before recording the variable whose value appears as missing, as it happens when a cow dies just after calving or when her calf is dead. The inclusion of this category, meaningless by itself, can however limit the bias in the estimation of the effects of the remaining categories.

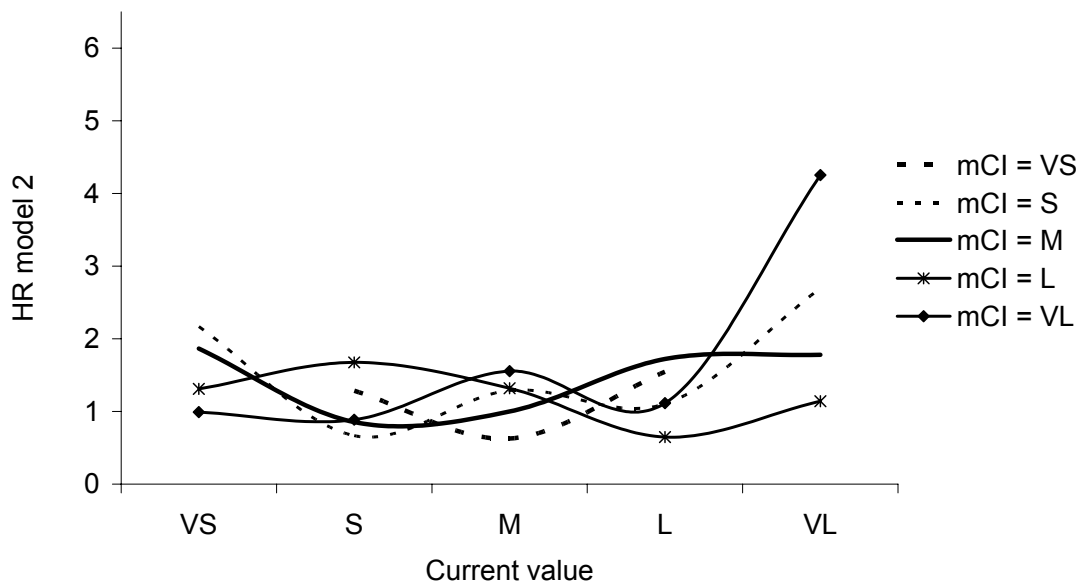
The inclusion of an interaction with parity indicated that the risk of primiparous cows seems higher than that of multiparous ones. Similar results were obtained in Avileña-Negra Iberica beef breed, where the risk of being culled seemed to be very important at the beginning of the productive life (Díaz et al., 2002). The risk of being

**Figure 6.** Effect of calving interval (CI) on culling expressed as relative hazard ratio (HR) (reference = M)

(a) Model 1



(b) Model 2, including an interaction with the mean of previous calvings (mCI)



(VS = Very small, S= Small, M = Medium, L = Large, VL = Very large)

culled of primiparous cows also increases with higher calving difficulty, larger calf birth weights, very small calf weight gains until weaning and very long calving intervals. However, these effects are not showed here because they did not reach statistical significance and further research would be needed to assess their relevance. In any case, the separation between primiparous and multiparous cows in this study is justified in order to limit bias in model 2 because the first calving has no interaction with previous calvings and so it can not be included in the model with other calvings that really have it.

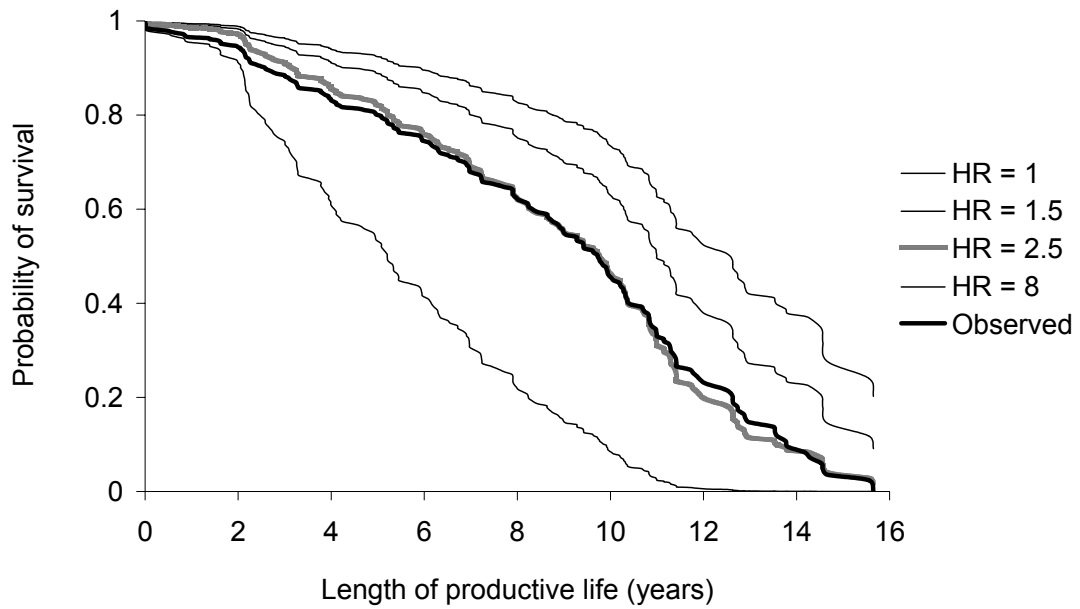
In summary, modelling time dependent variables in Model 1 allows the risk of culling to increase or decrease as the values of culling factors changed at each calving date. However, despite the limited size of our data set, Model 2 showed indications that the effect of the current calving values of culling factors also depend upon the mean of some performances of the previous calvings. These results would suggest that Model 1 would not be completely appropriate to correctly describe the culling process and it would be necessary to include the interaction between the current value of the culling factors and the mean of previous performances. This interaction effect could have an impact in the assessment of longevity values, for which models including it would deserve additional research.

### ***Useful measures of longevity***

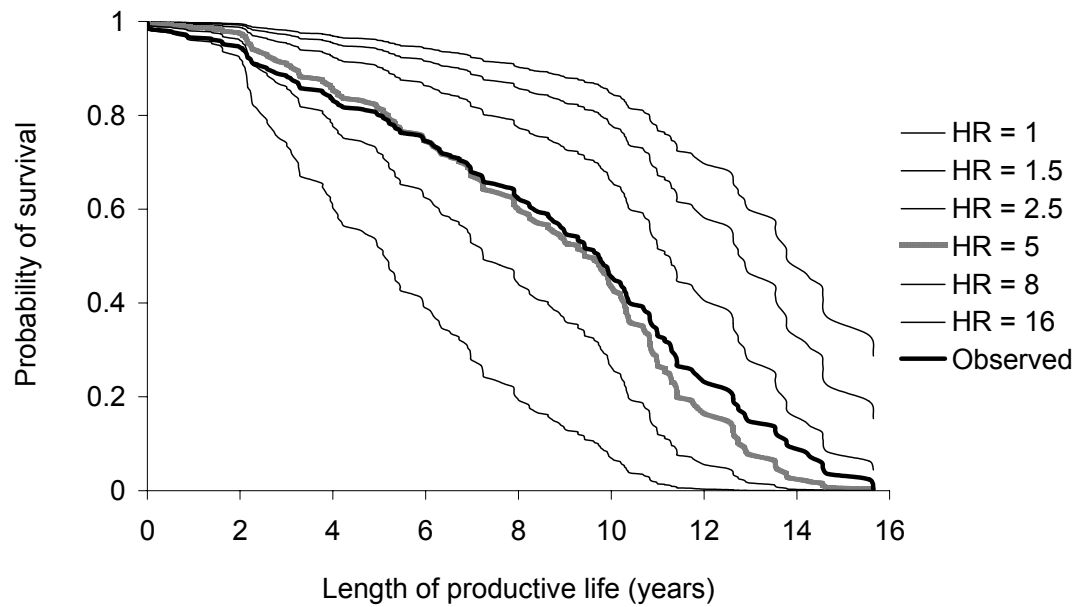
The effects of the culling factors have been presented so far using hazard ratios (HR). However, HR are relative measures that do not take into account the baseline hazard. The same HR (for example, an instantaneous culling risk multiplied by two) could be very important in species reared more intensively (like dairy cows and sows) because the baseline hazard is comparatively high, but not so important in species reared more extensively with a reduced baseline hazard. So, to obtain useful measures for future economic studies, survival functions with herd 1 were computed in both models assuming different hypothetical situations with constant hazard ratios (Figures 7a and 7b). These survival functions allow the computation of annual replacement rates (RR) and average calf weaning weights (AWW) weighted by the age structure in such a hypothetical herd. Table 4 shows that for the same HR, the corresponding values of RR and AWW in herd 1 for Model 1 and 2 are different. These differences exist because the baseline hazard in Model 1 is double than in Model 2. This is the direct consequence of the choice of reference animals: they are assumed to have intermediat

**Figure 7.** Expected survival functions estimated by model 1 (a) and model 2 (b).

(a) Model 1



(b) Model 2



current covariate values in Model 1, while in Model 2, they are also assumed to have intermediate values for previous calvings. Therefore, the hazard ratios cannot be directly compared between models.

The replacement rates for the reference values were 8.5 % and 7.7 % in Models 1 and 2, respectively, whereas the reference AWW were 236.4 kg and 236.7 kg. These would be the values of a homogeneous herd having intermediate values in all calvings. However, the animals in the herds analysed were under the influence of factors combined in many ways that increased culling compared to the reference situations (observed HR exceeded 2.5 in Model 1 and 5 in Model 2). These lower observed survival functions corresponded to a higher replacement rate (11.3%), and would decrease average weaning weights by only about 1 kg. However, the ideal situation with all animals in the herd at intermediate values is unrealistic. Any circumstance that might raise the reference HR by 8 in Model 1 or by 16 in Model 2 would increase the RR up to 18% but would decrease AWW by only about 3 kg. The above results show that in beef cattle reared extensively, a lower survival of the cows affects mainly the replacement rate. The effects on average weaning performances due to the changes in the structure of herd ages is not expected to be very important in beef cattle, although it could be more important in more intensive production systems where the proportion of animals producing at early ages is higher.

**Table 4.** Replacement rates (RR) and average weaning weights (AWW) in Herd 1 associated to different hypothetical relative hazard ratios (HR)

Model 1			Model 2		
HR	RR (%)	AWW (kg)	HR	RR (%)	AWW (kg)
1	8.5	236.4	1	7.7	236.7
1.5	9.5	236.1	1.5	8.3	236.5
2.5	11.2	235.6	2.5	9.4	236.2
			5	11.6	235.5
8	18.3	233.7	8	13.8	234.9
			16	18.8	233.5

## **Conclusions**

Age at first calving, calf birth weight, calving difficulty, calf weight gain until weaning and calving interval together with the herd-year effect were clearly related to the culling

process in the beef cattle population studied. There was an increased (probably voluntary) culling due to very small calf weight gain until weaning and very long calving intervals. On the other hand, there was a (probably involuntary) culling due to very high ages at first calving, higher calving difficulty and very large calf birth weights. These voluntary and involuntary culling decisions decreased the baseline survival function to the observed one. These lower than optimal survival characteristics would increase replacement rates but only slightly decrease the expected average performance of the herd.

Culling factors can be properly treated as time dependent variables. However, despite the fact that the performances in the current calving had a great influence on the risk of culling, their effect also depended upon the sequence of some performances at previous calvings. Thus, it is suggested that future genetic analysis of length of productive life in beef breeds should include time dependent variables, also taking into account the information from previous calvings.

### **Acknowledgements**

The suggestions of two anonymous referees contributed to deeply improve the manuscript. The “*Departament d’Agricultura, Ramaderia i Pesca de la Generalitat de Catalunya*” funded this work. J. Tarres was supported by a grant from the “*Ministerio de Educación, Cultura y Deporte*” of Spain. The contribution in collecting data of the farms *La Tria*, *La Torre* and *Borda la Ribera*, and of Chuck Simmons, who corrected the final English version of this manuscript, are greatly acknowledged.

### **References**

- Buxadé, C. 1997. *Vacuno de carne: aspectos claves*. Mundi-Prensa, Madrid.
- Casellas, J. and Piedrafita, J. 2002. Correction factors for weight productive traits up to weaning in the Bruna dels Pirineus beef cattle breed. *Animal Research* 51: 43-50.
- Cox, D. R. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society.*, Series B 34: 187-220.
- Díaz C., Chirinos Z., Moreno A., Carabaño M. J. 2002. Preliminary analysis of functional longevity in the Avileña Negra Ibérica beef cattle breed. *Proceedings of 7<sup>th</sup> World Congress on Genetics Applied to Livestock Production*, Montpellier, vol. 29, pp. 697-700.

- Ducrocq, V. P. 1997. Survival analysis, a statistical tool for length of productive life data. *48<sup>th</sup> Annual Meeting of the European Association for Animal Production*, Vienna.
- Ducrocq V. P., Quaas R. L., Pollak E. J., Casella G. 1988. Length of productive life of dairy cows. 1. Justification of a Weibull model. *Journal of Dairy Science*, 71: 3061-3070.
- Ducrocq V. P., Solkner. 1998. "The Survival Kit v3.0", a FORTRAN package for large analysis of survival data. *Proceedings of 6<sup>th</sup> World Congress on Genetics Applied to Livestock Production*, Armidale, vol. 27, pp. 447-450.
- Essl, A. 1998. Length of productive life in dairy cattle breeding: a review. *Livestock Production Science* 57: 79-89.
- Kalbfleisch, J. D. and Prentice, R. L. 1980. *The statistical analysis of failure time data*. Wiley, New York.
- Kaplan, E. L. and Meier, P. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53: 457-481.
- Luo, M. F., Boettcher, P. J., Schaeffer, L. R., Dekkers, J. C. M. 2002. Estimation of genetic parameters of calving ease in first and second parities of Canadian Holsteins using Bayesian methods. *Livestock Production Science* 74: 175-184.
- Rajala-Schultz, P. J. and Gröhn, Y. T. 2001. Comparison of economically optimized culling recommendations and current culling decisions of Finnish Ayrshire cows. *Preventive Veterinary Medicine* 49: 29-39.
- Tarrés J., Puig P., Piedrafita J. 2002. Non-genetic factors influencing productive life length and replacement rates in the Bruna dels Pirineus beef cattle breed. *Proceedings of 7<sup>th</sup> World Congress on Genetics Applied to Livestock Production*, Montpellier, vol. 29, pp. 693-696.
- Torrent, M. 1991. *La oveja y sus producciones*. AEDOS, Madrid.
- Veerkamp, R. F. , Brotherstone, S., Meuwissen T. H. E. 1999. Survival analysis using random regression models. *Interbull Bulletin* 21: 36-40.
- Yazdi, M. H., Rydhmer, L., Ringmar-Cederberg, E., Lundeheim, N. and Johansson, K. 2000. Genetic study of length of productive life in Swedish Landrace sows. *Livestock Production Science* 63: 255-264.

## **Capítol 3**

***GENETIC AND ENVIRONMENTAL FACTORS INFLUENCING  
MORTALITY UP TO WEANING OF BRUNA DELS PIRINEUS  
BEEF CALVES IN MOUNTAIN AREAS. A SURVIVAL ANALYSIS***

**J. Tarrés, J. Casellas, and J. Piedrafita**

*Grup de Recerca en Remugants, Departament de Ciència Animal i dels Aliments,  
Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*

***Journal of Animal Science 83, 543-551 (2005)***



## **Abstract**

Survival analysis techniques were used to analyze survival up to weaning of beef calves in the Pyrenean mountains areas of Catalonia, Spain. The Kaplan-Meier curve showed that the survival experience was not constant throughout the lactation period, as the mortality rate was more pronounced during the first month of life. The proportional hazards model analysis showed that several factors influenced the instantaneous mortality rate. The herd-year effect had the strongest influence. Calves born in the first part of the breeding season, from September to February, had the lowest mortality risk ( $P < 0.001$ ), showing that mortality risk increases as births accumulate. Calves from cows younger than 1,300 d of productive life had a higher risk of mortality ( $P < 0.05$ ). Non-assisted calvings presented the smallest risk of mortality, and it increased up to five times as birth became more difficult ( $P < 0.001$ ). This risk also tended to increase slightly when calf birth weight was small ( $P < 0.10$ ); for bigger calves, no increase of risk was detected, probably because calving difficulty was included in the model. These results suggest to improve the environment in the second part of the breeding period and to pay more attention to births from younger cows. The survival curve fitted a parametric piecewise exponential function very well, with cut points at 16 and 32 days. The lower risk corresponded to the period 33-180 days, the risk for the periods 17 to 32 days and 1 to 16 days being multiplied by 7 and 26, respectively. Confirming the robustness of the Cox model, the relative risks estimated for the different factors under this piecewise exponential model or a Weibull time-dependent model were similar to those reported above, as well as to those estimated under a frailty model, including the sire as a random effect. The modal estimates of sire variance under different baseline functions were close to 0.3, although the standard errors were very large. At weaning, the heritability estimate in the binary scale reached a value of only 0.037, because the survival at weaning was very high (96.9%) in this population. Nevertheless, in populations with a higher mortality, the inclusion of survival to weaning in the breeding objective might be justified. Overall, these results show that survival analysis is a powerful tool to analyze the mortality curve until weaning of beef calves.

**Key Words:** Beef Cattle, Heritability, Survival Analysis, Survival to Weaning

## **Introduction**

Improvement in cattle has traditionally focused on productive traits. However, modern cattle breeding programs should consider all traits of economic importance in order to optimize total genetic gain. The mortality of calves from birth to weaning reduces beef farm incomes and adds significantly to beef production costs (see Goyache et al., 2003, for a review). It is thus necessary to integrate calf survival traits into the definition of an overall breeding objective in beef cattle.

Despite its economic importance for suckling cow production systems, phenotypic and genetic studies on calf survival at weaning are scarce and little information is available (Cundiff et al., 1986; Ray et al., 1989; Goyache et al., 2003). Until now, calf survival has usually been studied as a dichotomous variable at a given age. However, the use of a continuous measure as failure time is preferable to analyzing it with regression models since it makes use of all the information available more efficiently by not restricting observations to an arbitrarily defined point (Ducrocq, 1997). The problem of working with these continuous measures is the presence of live animals at the completion of the study that have inexact information of their failure time. Records from these animals are classified as censored. Survival analysis is the method of choice to analyze the failure time data allowing the inclusion of censored and uncensored (dead animals) records. The aim of this paper is to analyze calf survival using survival analysis techniques. The general strategy for the analysis of survival data in beef calves is: description of the observed survival function by the Kaplan-Meier method, estimation of fixed effects with a semiparametric proportional hazards model and validation of the baseline hazard function, and finally the estimation of the sire variance component for hazard and the heritability of survival in the binomial scale.

## **Material and Methods**

### ***Data***

Bruna dels Pirineus is a beef breed selected from the old Brown Swiss (derived from the Canton Schwyz), similar to the American Braunvieh, located in the eastern part of the Pyrenees. The herds are located in the Pyrenean mountain areas of Catalonia (Spain). The production system includes a stay in valleys close to villages from

October-November to June, when most of the calvings occur, and afterwards cows and calves are taken to the mountains to graze alpine pastures.

Data were recorded between 1994 and 2002 in three breeding herds participating in the Yield Recording Scheme of the breed, and included data from 3,335 beef calves. The pedigree of the calves was only partly known, as 86 bulls sired 2,578 calves, the rest of the sires being unknown. All dams were known. The survival time was estimated as the difference between the date of death and the date of birth. When the date of death was unknown, we assumed that the calf record was censored at the date of weaning (180 d), and any known dates of death after weaning were not used in the analysis. After editing and excluding twin births, our database included data of survival time of 2,504 calves, with 68 complete records (2.7%) and 2,436 censored records (97.3%).

### **Survival Analysis**

Our survival analysis followed the general strategy proposed by Ducrocq (1997). In the preliminary analysis, the observed distribution of survival time was estimated empirically by the Kaplan-Meier method (Kaplan and Meier, 1958). This observed distribution was the result of the influence of different factors involved in the survival process. To investigate the influence of these factors on the instantaneous death rate ( $h(t; \mathbf{X})$ , the hazard) the following semiparametric proportional hazards model (Cox, 1972) was assumed:

$$h(t; \mathbf{X}) = h_0(t) \exp[HY + MC_m + LPL_j + CD_k + BW_i] \quad [1]$$

The first term of the right hand side is the unknown baseline hazard function  $h_0(t)$ . The second term (between brackets) was dependent upon a subset of explanatory variables, which we call survival factors. Herd×year of calving ( $HY$ ), month of calving ( $MC$ ), cow's length of productive life at calving ( $LPL$ ), calving difficulty ( $CD$ ) and birth weight ( $BW$ ) were treated as time independent variables and included sequentially from left to right in Models 1 to 4. Hypothesis testing was performed via likelihood ratio tests. The *sequential* test included the effects in the model in sequential order. The *last* test compared the full model with models excluding one effect at a time (this was done for

each effect separately). These and the subsequent computations were performed by means of the Survival Kit package (Ducrocq and Sölkner, 1998).

**Table 1.** Number of uncensored and total animals in each level of each survival factor

Survival factor	Classes	Uncensored	Total
Month of calving	September to February	22	1369
	March to August	46	1135
Productive life <sup>a</sup>	< 1300 d	50	1397
	> 1300 d	18	1107
Calving difficulty	1	36	1935
	2	10	136
	3	8	116
	Missing	14	317
Calf birth weight	Small	20	522
	Medium/Large	22	1074
	Missing	26	908
Total		68	2504

<sup>a</sup>Cow's length of productive life at calving.

Because the form of the phenotypic relationship between survival factors and the hazard was not known, all variables included in the models were categorized (Table 1). This way, no assumption about the nature of this relationship has to be made (Ducrocq, 1997). Herd×year effect (HY) was treated as a fixed variable. The inclusion of an interaction term between herd and year of calving would permit precise modeling of changes in the mortality process within a farm over time, similar to that observed by Ducrocq (1997) for longevity in dairy cattle. After a preliminary study, classes with a similar effect on the risk of mortality were grouped. Month of calving had two classes with a similar number of observations (September to February and March to August). Length of productive life of the cow at calving, i.e., the difference between the age of the cow at calving and the age at first calving, had two classes, 1 to 1,300 d (cows up to four calvings) and productive life longer than 1,300 d. Calving difficulty had four classes: calving without assistance, calving slightly assisted by the farmer, calving strongly assisted by the farmer or the veterinary practitioner, and missing values. Finally, calf birth weight was grouped into three classes: small weights (up to 42.9 kg), medium to large weights, and missing values.

The proportional hazards assumption was tested by extending the proportional hazards model with time-dependent factors (i.e., interaction terms between the time-independent (or proportional) factor and a function of time). The proportional hazards assumption was checked via likelihood ratio tests (LRT) comparing the proportional model with models including one time dependent factor at a time.

**Validation of the baseline hazard function**

The attractive feature of the semiparametric proportional hazard model is that it permits the estimation of the regression coefficients in  $\beta$  without making any assumption about the form of the baseline distribution function. However, the estimation of fixed and random effects and of genetic parameters is much less demanding of computing resources with a parametric model than with a semiparametric model (Yazdi et al., 2002). A classical approach to choose an adequate form for the distribution function of a random variable is to estimate its empirical survival function (Kaplan-Meyer estimate) and to compare its shape to that of known parametric distributions (Ducrocq, 1997). In animal breeding, the Weibull model has been the parametric proportional hazards model of choice when longevity is analyzed. The Weibull model is a very flexible parametric model, in particular, when time-dependent covariates are included in the model (Yazdi et al., 2002).

The Weibull hazard distribution is  $h_0(t) = \lambda\rho(\lambda t)^{\rho-1}$ , the Weibull baseline survival function being  $S_0(t) = \exp[-(\lambda t)^\rho]$ . Then, a plot of  $\ln(-\ln S_0(t))$  or  $\ln(-\ln \hat{S}_{KM}(t))$  against  $\ln t$  gives a straight line:  $\ln(-\ln S_0(t)) = \rho \ln \lambda + \rho \ln t$  (Ducrocq, 1997). When  $\rho = 1$ , the Weibull hazard function is constant and is reduced to an exponential hazard function,  $h_0(t) = \lambda$ , the exponential baseline survival function being  $S_0(t) = \exp[-\lambda t]$ . Similarly, a plot of  $\ln S_0(t)$  against  $t$  gives a straight line as  $\ln S_0(t) = -\lambda t$  and thus, the graphical test indicating whether an exponential model is suitable consists of checking that a plot of  $\ln \hat{S}_{KM}(t)$  against  $t$  gives a straight line.

The inclusion of time-dependent covariates  $P_c$  converts the exponential hazard function into a piecewise exponential function:

$$h_0(t) = \begin{cases} \lambda_1 = \lambda \exp(P_1) & \text{if } t \in (0, c_1) \\ \lambda_2 = \lambda \exp(P_2) & \text{if } t \in (c_1, c_2) \\ \vdots & \\ \lambda_{k+1} = \lambda \exp(P_{k+1}) & \text{if } t \geq c_k \end{cases} \quad [2]$$

These formulas assume that the hazard of an individual is constant within each period of time but is different for the same individual between the  $c = 1, 2, \dots, k + 1$  different periods, corresponding to  $k$  cut points. Furthermore, the survival function associated with any hazard function is the exponential of the negative cumulative hazard function,

$$S(t) = \exp\left[-\int_0^t h(u) du\right], \text{ and the piecewise exponential function becomes:}$$

$$S_0(t) = \begin{cases} \exp\left[-\int_0^t \lambda_1 dt\right] = \exp(-\lambda_1 t) & \text{for } t \in (0, c_1) \\ \exp\left[-\int_0^{c_1} \lambda_1 dt - \int_{c_1}^t \lambda_2 dt\right] = \exp(-\lambda_1 c_1 - \lambda_2 (t - c_1)) & \text{for } t \in (c_1, c_2) \\ \vdots & \\ \exp\left[-\int_0^{c_1} \lambda_1 dt - \int_{c_1}^{c_2} \lambda_2 dt - \dots - \int_{c_k}^t \lambda_{k+1} dt\right] = \exp(-\lambda_1 c_1 - \dots - \lambda_{k+1} (t - c_k)) & \text{for } t \geq c_k \end{cases}$$

There are  $k + 1$  different formulas for piecewise exponential functions. These formulas can be summarized as:

$$S_0(t) = \exp\left(-\left[\sum_{j=1}^l \lambda_j (c_j - c_{j-1})\right] - \lambda_{l+1} (t - c_l)\right) \quad \text{if } t \in (c_l, c_{l+1}) \quad [3]$$

One property of the piecewise exponential function is that a plot of  $\ln S_0(t)$  against  $t$  yields  $c = 1, 2, \dots, k + 1$  different straight lines:

$$\ln(S_0(t)) = \begin{cases} -\lambda_1 t & \text{for } t \in (0, c_1) \\ -\lambda_1 c_1 - \lambda_2 (t - c_1) & \text{for } t \in (c_1, c_2) \\ \vdots & \\ -\lambda_1 c_1 - \lambda_2 (c_2 - c_1) - \dots - \lambda_{k+1} (t - c_k) & \text{for } t \geq c_k \end{cases} \quad [4]$$

This property suggests a graphical test for the piecewise exponential model that consists of checking that a plot of  $\ln \hat{S}_{KM}(t)$  against  $t$  yields  $c = 1, 2, \dots, k + 1$  different straight lines. The slopes of these lines are the hazards in each period of time

$(\lambda_1, \lambda_2, \dots, \lambda_{k+1})$ . To choose the cut points  $(c_1, c_2, \dots, c_k)$  of the piecewise function we fit splines in the regression of the log of the KM survival function on time. To do that, all possible combinations of  $1, 2, \dots, k$  cut points in the time space (180 d) were explored, avoiding combinations of adjacent days.

As before, the inclusion of time-dependent covariates  $P_c$  converts the Weibull hazard function into a Weibull time-dependent (WeiTD) hazard function:

$$h_0(t) = \begin{cases} \lambda_1(t) = \lambda \rho (\lambda t)^{\rho-1} \exp(P_1) & \text{if } t \in (0, c_1) \\ \lambda_2(t) = \lambda \rho (\lambda t)^{\rho-1} \exp(P_2) & \text{if } t \in (c_1, c_2) \\ \vdots & \\ \lambda_{k+1}(t) = \lambda \rho (\lambda t)^{\rho-1} \exp(P_{k+1}) & \text{if } t \geq c_k \end{cases} \quad [5]$$

Again, the survival function associated is the exponential of the negative cumulative hazard function,  $S(t) = \exp\left[-\int_0^t h(u) du\right]$  (formulas not shown). The Weibull time-dependent function has more degrees of freedom than its embedded distributions (Weibull, exponential and piecewise) and thus improves the likelihood of the model. This hypothesis was tested via likelihood ratio tests (LRT). In our analysis, a comparison between a Weibull time-dependent model and its embedded models was performed. The LRT under the null hypothesis was defined as:

$$LRT = [-2 \log L_{imbedded} + 2 \log L_{weiTD}] \quad \text{asympt} \sim \chi_q^2$$

In this case, the LRT statistic has 1, k and k + 1 degrees of freedom for the piecewise, Weibull and exponential models, respectively.

### **Random effects**

The proportional hazards model can be extended to a frailty model with the addition of a random sire effect  $\mathbf{s}$ :

$$h(t; \mathbf{X}, \mathbf{Z}) = h_0(t) \exp[\mathbf{X}\boldsymbol{\beta}] \exp[\mathbf{Z}\mathbf{s}] \quad [6]$$

The vector  $\mathbf{s}$  of additive sire genetic effects, under polygenic inheritance, was assumed to follow a multivariate normal distribution  $\mathbf{s} \sim MVN(\mathbf{0}, \mathbf{A}\sigma_s^2)$ , where  $\mathbf{A}$  is the additive genetic relationship matrix among sires. The sire variance ( $\sigma_s^2$ ) was estimated using the Bayesian approach described in Ducrocq and Casella (1996). An obvious

point estimate of the sire variance is the mode of this approximate marginal posterior density  $\hat{\sigma}_s^2$ . The general characteristics of the distribution through the computation of its first three moments by unidimensional numerical integration based on Gauss-Hermite quadrature have also been computed (Ducrocq and Casella, 1996).

### **Heritability**

The heritability on the binary scale ( $h_{bin}^2$ ) is given by the formula

$$h_{bin}^2 \approx \frac{4 \sigma_s^2}{\sigma_s^2 + \frac{1 - S_0(t)}{[\ln(S_0(t))]^2 S_0(t)}} \quad [7]$$

similar to that reported by Yazdi et al. (2002). Here, the mode of the sire variance  $\hat{\sigma}_s^2$  is used. The heritability changes along time with the value of the survival function. Its derivation is presented in the Appendix.

## **Results**

### ***Observed calf survival***

The observed survival function estimated by the Kaplan-Meier method is presented in Figure 1. The decrease in the observed probability of survival was more pronounced until the end of the first month (97.5%) and after that point the decrease was much lower until weaning (96.9%).

### ***Survival Factors***

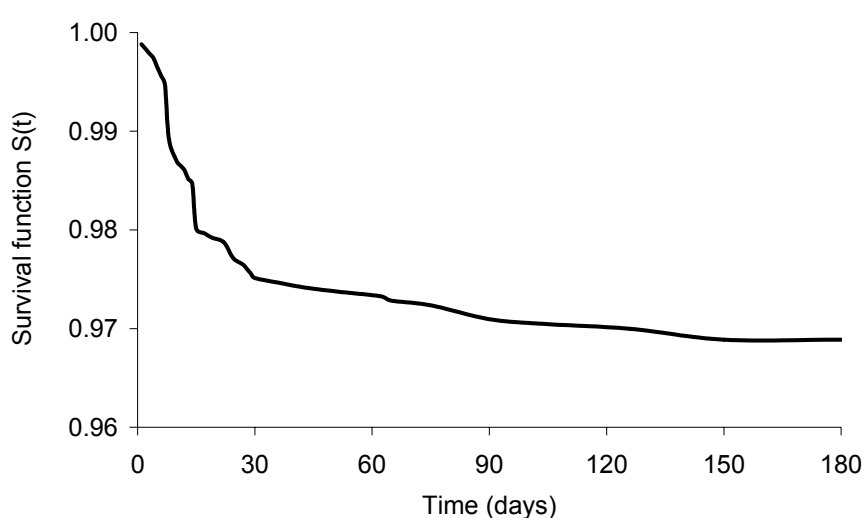
All analyzed factors (i.e., herd-year, month of calving, cow's length of productive life at calving, calving difficulty and calf birth weight) were significantly related to the risk of mortality when they were entered sequentially in the model or excluded from the full model, one at a time (Table 2). In this last test, however, the significance of the cow's length of productive life was lower probably due to its correlation with the other factors.

These survival factors were assumed to be time-independent (or proportional). To test the proportional hazards assumption, time-dependent factors (i.e., interaction



terms between the time-independent factors and a function of time (changing at 16 and 32 d)) were defined. The inclusion of these interaction terms did not increase the likelihood significantly in any of the models analyzed (results not shown in tables), and, thus, the proportionality hypothesis was not rejected. This implies that the estimated  $\beta$ -coefficients for each proportional factor are equivalent to the effect of the variable throughout the period analyzed (Allison, 1995).

**Figure 1.** Observed survival function (Kaplan-Meier estimate) of beef calves from birth to weaning.



**Table 2.** Likelihood statistics for contrasts of survival factors

Factor	DF	Sequential <sup>a</sup>	Last <sup>b</sup>
Herd-year	18	45.65 ***	40.58 ***
Month of calving	1	12.51 ***	12.23 ***
Productive life <sup>c</sup>	1	8.61 **	3.60 †
Calving difficulty	3	18.32 ***	21.36 ***
Calf birth weight	2	6.21 *	6.21 *
Total	25	91.3 ***	

( † p< 0.10; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001; NS non-significant)

<sup>a</sup> Factors were included in the model in sequential order, the numbers expressing the increase of likelihood after including the factor (Sequential).

<sup>b</sup> The full model was compared with models excluding one effect at a time, the numbers standing for the loss of likelihood after excluding the effect (Last).

<sup>c</sup> Cow's length of productive life at calving.

### **Hazard ratios of survival factors**

Herd-year effect had the strongest influence on calf risk of mortality (Table 2). The mortality rate could change every year due to several sources of variation, such as natural forage availability, changes in herd management practices and other environmental circumstances. The estimates of the corresponding hazard ratios were not shown as they did not follow any consistent trend.

**Table 3.** Hazard ratios of survival factors included in the model in sequential order

Survival factor	Classes	Model 1	Model 2	Model 3	Model 4
Month of calving	September to February	1	1	1	1
	March to August	2.678***	2.536***	2.588***	2.570***
Productive life <sup>c</sup>	< 1,300 d		2.214**	2.000*	1.682 <sup>†</sup>
	> 1,300 d		1	1	1
Calving difficulty	1			1	1
	2			3.170**	2.776*
	3			5.305***	6.911***
	Missing			2.845	4.832
Calf birth weight	Small				1.884 <sup>†</sup>
	Medium/Large				1
	Missing				0.684

(<sup>†</sup>p< 0.10; \* p<0.05; \*\* p<0.01; \*\*\* p<0.001; NS non-significant)

<sup>a</sup>Cow's length of productive life at calving.

Calves born from September to February had the lowest mortality risk, increasing from March to August ( $P < 0.001$ ) (Table 3). The calving season starts after the return of the cows from the mountains, in September or October, and by the beginning of March half of the cows had calved. That is, the risk of mortality increases as does the cumulated distribution of calvings, probably due to a higher competition for resources and a higher risk of exposure to infectious diseases.

Cows older than 1,300 d of productive life bore calves with a lower risk of mortality than younger cows (Table 3). This increased risk was observed until the fourth calving. However, the differences in risk of death between progeny from young and old cows were lower when calving difficulty and birth weight entered into the

model. This is probably due to the correlation between the variables involved, previously reported in this population by Casellas and Piedrafita (2002).

Non-assisted calvings presented the minimum risk of mortality, and the risk increased as birth became more difficult (Table 3). When assistance was very strong, calf mortality risk was more than five times that of calves born in a non-assisted parturition ( $P < 0.001$ ). Missing values probably corresponded to animals that were at a high risk of death before recording the variable whose value appears as missing, as happens when a calf dies just after birth. The inclusion of this category, however, might limit the bias in the estimation of the effects of the remaining categories.

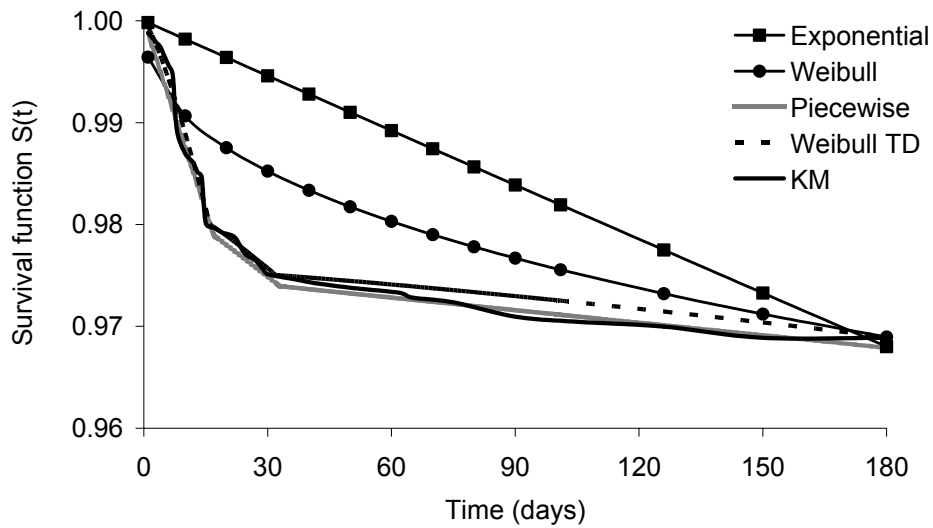
Calves with medium and large birth weights had the lowest risk of death (Table 3). This hazard tended to increase slightly when calf birth weights were small ( $P < 0.10$ ). Future studies including a larger data set will probably help to clarify the true effect of the birth weight of a calf on its future survival.

### ***Validation of a Baseline Hazard Function***

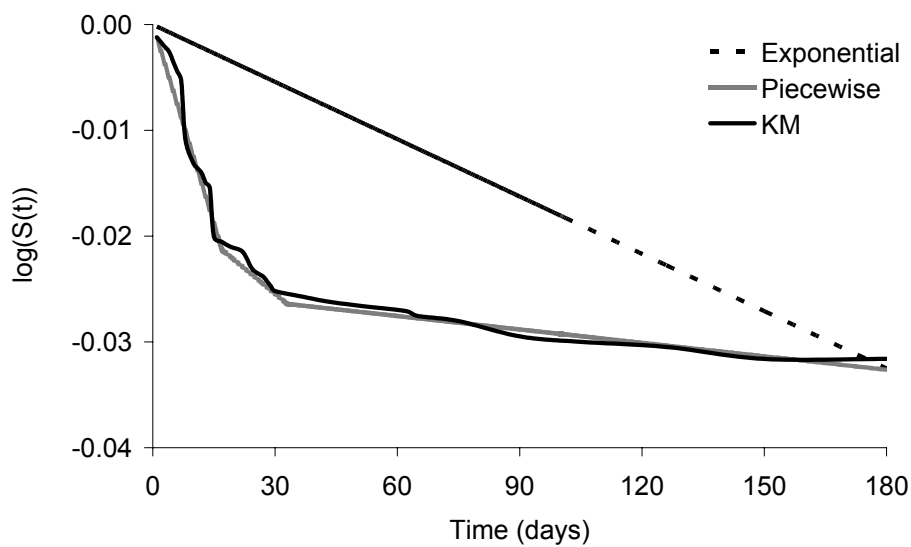
The Kaplan-Meier survival function may be approximated by a parametric function. Figure 2 presents plots for the Kaplan-Meier estimate of the survival function in comparison to plots of four parametric functions estimated in a model with no covariates from our data set. At first sight, both piecewise exponential and Weibull time-dependent functions showed a close fit to the observed (Kaplan-Meier estimate) survival function. The plots presented in Figures 3a and 3b confirm this assertion and show that the observed function neither fits the Weibull nor the exponential time independent functions. However, the observed curve was well approximated both by a piecewise exponential and a Weibull time dependent distribution with cut points at 16 and 32 d. Furthermore, as the likelihoods of both piecewise and Weibull time-dependent survival functions were not statistically different for all models studied (results not shown in tables), the piecewise exponential model could be the function of choice to describe the mortality process of beef calves from birth to weaning.

For the piecewise exponential function, the slopes of the three lines were  $\lambda_1 = 0.001226$  during the first two weeks,  $\lambda_2 = 0.000332$  from 16 to 32 d and  $\lambda_3 = 0.000046$  after the first month until weaning (Figure 3b). Thus, the instantaneous mortality rate (risk) was more pronounced (26 times) during the first two weeks of life, diminishing from 16 to 32 d, and after that was kept much lower until weaning, confirming what was observed by simple inspection of the K-M curve.

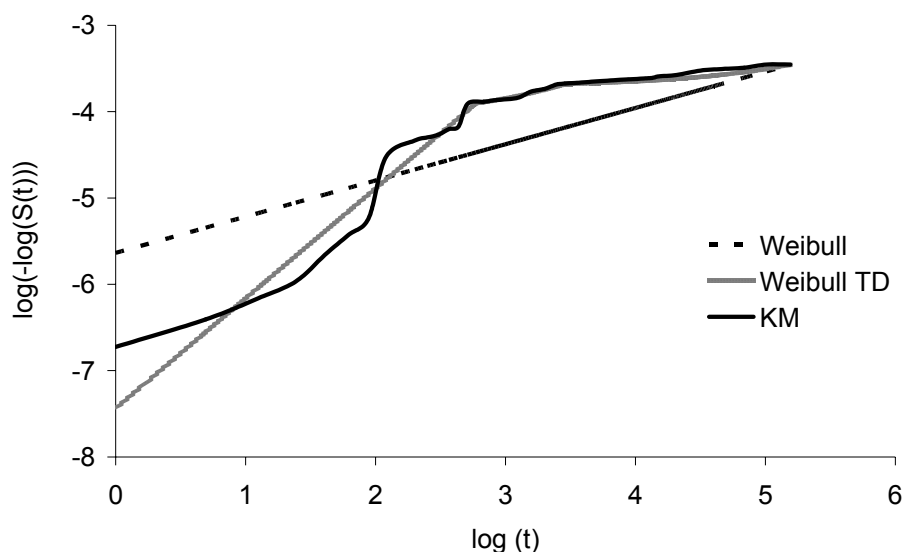
**Figure 2.** Comparison of the Kaplan-Meier estimate (KM) with four parametric estimates of the survival function: exponential, piecewise exponential, Weibull and Weibull time-dependent (WeibullTD).



**Figure 3a.** Validation of exponential and exponential time dependent (Piecewise) functions compared to the observed (KM) function.



**Figure 3b.** Validation of Weibull and Weibull time dependent (WeibullTD) functions compared to the observed (KM) function.



### Frailty models

Similar estimates of hazard ratios for survival factors were obtained under a frailty model assuming different baseline hazard functions, with the exception of the exponential model (Table 4). The estimates were also very similar to those obtained under the fixed model.

The modal estimates of sire variance were close to 0.3 when all survival factors were included in the model and the validated baseline functions (piecewise, WeiTD) and the Cox model were assumed (Table 4). Other distributions that do not fit the data well may (exponential) or may not (Weibull) affect the estimate of the sire variance. This shows that both fixed effects and sire variance estimates were, in general, robust for different baselines. However, the standard error of the estimates was very large in all cases (about 0.38 to 0.47), and particularly under a Cox model (0.48), for which the null value was always included in the 95% high posterior density regions.

On the other hand, the estimates of the random effect (sire variance) decreased with the number of survival factors included in the model (results not shown in tables). For Models 2 and 3, the sire variance estimates for the different baselines assumed were in the 0.52 to 0.54 and 0.46 to 0.51 ranges, respectively, whereas for Model 4 the estimates oscillated between 0.28 and 0.31.

**Table 4.** Baseline parameters, hazard ratios of fixed effects and sire variance estimates of calf survival under different baselines

Factor	Classes	Exponential	Weibull	Piecewise	WeiTD <sup>a</sup>	Cox
Rho			0.443478		1.32342	
Lambda		0.0001369	0.002629	0.0009923	0.0004034	
Period	<15 d			1	1	
	16 to 31 d			0.271	0.178	
	> 32 d			0.038	0.016	
Month of calving	September to February	1	1	1	1	1
	March to August	3.017	2.664	2.598	2.636	2.558
Productive life <sup>b</sup>	< 1,300 d	1.792	1.778	1.788	1.795	1.790
	> 1,300 d	1	1	1	1	1
Calving difficulty	1	1	1	1	1	1
	2	2.969	2.818	2.751	2.766	2.705
	3	7.381	6.492	6.412	6.557	6.372
	Missing	23.543	17.475	18.176	19.331	18.902
Birth weigh	Small	1.856	1.796	1.782	1.785	1.782
	Medium / Large	1	1	1	1	1
	Missing	0.961	0.817	0.832	0.855	0.856
Sire variance	Mode	0.51518	0.29176	0.27876	0.31101	0.29960
	Mean	0.69377	0.56107	0.55291	0.58380	0.63004
	SD	0.43154	0.37554	0.38553	0.38629	0.46574
	Skewness	1.539	0.702	0.953	0.738	0.974

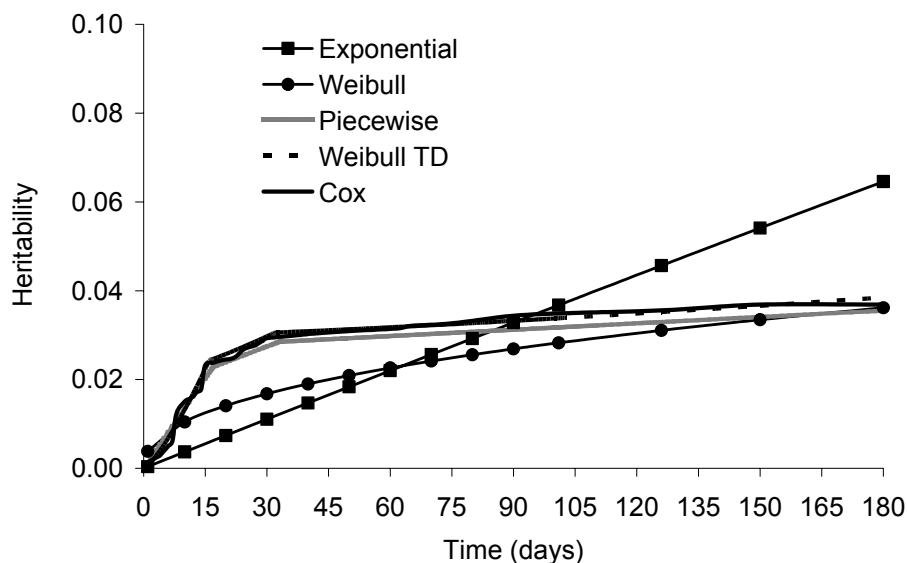
<sup>a</sup>Weibull time dependent base line.

<sup>b</sup>Cow's length of productive life at calving.

## Heritability

Heritability is not constant along the period studied, as it depends upon the survival at each time point and the modal estimate of the sire variance. Assuming a sire variance of 0.3 and baseline values close to the KM estimates of the survival curve, heritability was always low, particularly during the first days of life. At 15 d, heritability is 0.023 and only reaches a value of 0.037 at weaning (180 d), assuming a piecewise exponential model (Figure 4).

**Figure 4.** Estimated heritabilities over time assuming different baseline functions



## **Discussion**

Our estimate of calf death loss by weaning (3.1%) is lower than most other estimates reported in beef cattle populations. Ray et al. (1989) observed an average death loss of 3.0%, whereas Fagerling et al. (1968) found a yearly average loss of 9.2% for a Hereford herd in Wyoming. Ensminger (1977) and Cundiff et al. (1986) gave estimates of average death loss from birth to weaning for beef cattle of 6.0% and 8.0%, respectively. Morris et al. (1986) reported even much higher death losses from birth to weaning in primiparous 2-yr-old dams in research conducted in New Zealand (13.4% and 14.7% in two separate studies).

According to Ray et al. (1989) the two major components of per cent calf crop weaned are birth rate and calf survival. Thus, calf survival is a parameter conditional on a calf born alive. In our case evidence exists that some of the animals were born alive although an early death some hours after calving led the farmer to record this calf as dead. The consequence is that our survival would be an upwards estimate of the true survival.

The survival experience was not constant over the lactation period. The mortality rate was more pronounced during the first two weeks of life, diminishing from 16 to 32 d, and after that was kept fairly constant until weaning. This observation is in agreement with several previous reports. Meijering (1984), Bellows et al. (1987), and

Patterson et al. (1987) reported in beef-calf surveys that 75% of mortality of all causes occurred by 7 d of age. Furthermore, Thurmond (1986) found in dairy bull and heifer calves that the majority of deaths occurred between 9 and 16 d of age. These different mortality rates at different periods caused the Weibull distribution to fail in the validation of the baseline distribution, and a time-dependent distribution was needed. In this paper, the piecewise exponential function and a simple graphical method to validate its adequacy is proposed for calf survival analysis.

Several survival factors had an influence on calf survival. Herd-year effect had the strongest influence (Table 3), as was reported previously (Ray et al., 1989; Goyache et al., 2003). Calf survival also depended upon the season of calving. A similar phenomenon has been observed in Ripollésa Lambs (J. Casellas, personal communication). Furthermore, Cubas et al. (1989) and Goyache et al. (2003) also included a calving season effect with two levels (from January 1 to June 30 and from July 1 to December 31) in the analysis of calf survival.

Calves born from cows younger than 1,300 d of productive life had a higher risk of death than calves born from the older cows. Greater death losses among calves from primiparous cows is a common observation that has been previously reported in Hereford, Angus, Charolais, Brahman and crossbred cattle (see Ray et al., 1989, and the review by Cundiff et al., 1986). This fact could be explained, among other reasons, by an improvement of the maternal behavior as the cow matures.

Non-assisted calvings showed the minimum instantaneous mortality rate, with the rate increasing as birth became more difficult. Cundiff et al. (1986) reviewed several reports that have shown that calving difficulty has a significant effect on calf survival (Laster and Gregory, 1973; Smith et al., 1976; Gregory et al., 1978; Menissier et al., 1981). More recently, dystocia has been associated with calf mortality because of the possibility of uterine fluid inhalation (Bellows et al., 1987). However, a proportion of this effect is genetic, as perinatal survival has a high positive genetic correlation with calving ease (Cubas et al., 1989). In view of this high correlation, breeders interested in the improvement of calf survival rates normally have been recommended to use a correlated response to direct selection for calving ease (Cubas et al., 1989).

Calves with medium and large birth weights had the lowest instantaneous mortality rate. This rate tended to increase slightly when calf birth weights were small. Similar results have been reported in the literature (Kulkarni et al., 1994; review of Moore et al., 2002), although there are some other problems associated with lighter birth weights. Bellows et al. (1987) reported that 20% of abnormal beef calves



(determined by necropsy) had birth weights at least 3.6 kg lighter than the average normal birth weight of 35 kg. Earlier onset of calf diarrhea was seen with the lightest weight dairy heifer calves (Pare et al., 1993). Fallon et al. (1987) reported that lighter Friesian bull-calves (< 40 kg) were at higher risk for diarrhea and mortality than calves greater than 39 kg. A lower probability of survival of small birth weights was also reported in piglets. Casellas et al. (2004) described a direct effect of birth weight on piglet survival justifying that small piglets need to spend more energy for thermoregulation purposes and generally suffer a delay in the first colostrum intake and a variable degree of physiological immaturity. On the other hand, Moore et al. (2002) also reported that heavier calves had higher mortality because, as suggested by Meijering (1984), they may be responsible for dystocia due to feto-pelvic incompatibility. This effect was not detected in our study probably because of the inclusion of calving difficulty in the model. In future studies, calving difficulty might be analyzed jointly with other factors like birth weight, weaning weight and survival rate in a multivariate analysis taking advantage of the genetic correlations among these traits.

The genetic random effect was estimated assuming a sire model. A sire-maternal grandsire model was also explored. Variance components estimated with both models were very similar. The sire-maternal grandsire model does not distinguish between direct and maternal effects that usually are very important in pre-weaning traits of beef cattle (see for example Meijering, 1984; Cubas et al., 1989; Goyache et al. 2003). The ideal model of analysis would have been an animal model with maternal effects, but the structure of our data set did not allow us to develop such a model.

The sire variance estimated was around 0.3 for calf survival. This variance, estimated by assuming a proportional hazards model, could not be compared to the literature because published papers about calf survival estimated genetic variances assuming a threshold model. Until now, survival analysis techniques were only used to analyze cow longevity. Rogers et al. (2004) estimated sire variances for beef cows longevity (between 0.028 and 0.037). In this sense, our sire variance for calf survival was very much higher than the cow longevity variances. However, cow longevity and calf survival are different characters and thus may not be comparable.

In spite of this relatively high genetic variation, the corresponding heritability estimate for calf survival was low (0.037) at weaning (piecewise exponential model). This can be explained by the fact that heritability in the binary scale also depends on the survival at a certain age, which in our case is very high at weaning (96.9%). For some other estimates of the survival at weaning published in the literature (e.g., 0.95,

0.90 and 0.85) the heritabilities would be 0.059, 0.116 and 0.176, respectively. These estimates are in agreement with those reported in the literature in which a threshold model was assumed. Cundiff et al. (1986), using a multibreed population, estimated a heritability for the direct effect of survival from birth to weaning of 0.07 within sire breeds and of 0.11 for the total population. Goyache et al. (2003) estimated heritabilities considering the variables as calf traits ranging from 0.057 for late mortality (survival from 72 h to weaning) to 0.106 for weaning survival (from birth to weaning).

## **Conclusions**

Survival analysis can be a powerful tool to investigate and assess the survival up to weaning in beef calves. To improve calf survival two types of actions could be undertaken. On the environmental side, special attention must be paid to calves born from February to June and from young cows and especially primiparous cows. On the genetic side, the heritability found for calf survival in our data set was low. Nevertheless, with higher mortality rates, if the sire variance component for hazard could be confirmed, the inclusion of calf survival in the breeding objective would be justified. Furthermore, genetic improvement of calf survival also might take advantage of the correlated response to direct selection for reducing calving difficulty and maintaining intermediate birth weights.

## **Acknowledgements**

Financial support was provided by the “*Servei de Producció Ramadera, Departament d’Agricultura, Ramaderia i Pesca, Generalitat de Catalunya*”. J. Tarrés was supported by a grant from the “*Ministerio de Educación, Cultura y Deporte*” of Spain. The authors are indebt to the farms La Tria, Casa Vallnova and Mas Fogonella for contributing field data and to V. Ducrocq for his suggestions. The English revision by C. Simmons is also acknowledged.

## **References**

Allison, P. D. 1995. Survival Analysis Using the SAS® System. A Practical Guide. SAS Institute Inc., Cary, North Carolina.

- Bellows, R. A., D. J. Patterson, P. J. Burfening, and D. A. Phelps. 1987. Occurrence of neonatal and postnatal mortality in range beef cattle. II. Factors contributing to calf death. *Theriogenology* 28:573-586.
- Casellas J., and J. Piedrafita. 2002. Correction factors for weight productive traits up to weaning in the Bruna dels Pirineus beef cattle breed. *Anim. Res.* 51:43-50.
- Casellas J., J. L. Noguera, L. Varona, A. Sánchez, M. Arqué, and J. Piedrafita. 2004. Viability of Iberian × Meishan F<sub>2</sub> newborn pigs. II. Survival analysis up to weaning. *J. Anim. Sci.* 82:1925-1930.
- Cox, D. R., 1972. Regression models and life tables (with discussion). *J. Royal Stat. Soc. Series B* 34:187-220.
- Cubas, A. C., P. J. Berger, and M. H. Healey. 1989. Genetic parameters for calving ease and survival at birth in Angus field data. *J. Anim. Sci.* 69:3952-3958.
- Cundiff, L. V., M. D. MacNeil, K. E. Gregory, and R. M. Koch. 1986. Between and within-breed genetic analysis of calving traits and survival to weaning in beef cattle. *J. Anim. Sci.* 63:27-33.
- Ducrocq, V. P. 1997. Survival analysis, a statistical tool for length of productive life data. 48<sup>th</sup> Annual Meeting of the European Association for Animal Production, Vienna. [http://www.nas.boku.ac.at/fileadmin/\\_/H930-agrar/H932-NUWI/Software/Survival/e97surv.doc](http://www.nas.boku.ac.at/fileadmin/_/H930-agrar/H932-NUWI/Software/Survival/e97surv.doc)
- Ducrocq, V. P., and G. Casella. 1996. A Bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 28:505-529.
- Ducrocq, V. P., and J. Sölkner. 1998. The Survival Kit v3.12", a FORTRAN package for large analysis of survival data. *Proc. 6<sup>th</sup> World Cong. Genet. Applied Livest. Prod., Armidale, Australia* 27:447-450.
- Ensminger, M. E. 1977. *Animal Science*. The Interstate Printers and Publishers Co., Danville, IL.
- Fagerling, P. T., J. S. Brinks, and H. H. Stonaker. 1968. Environmental effects on calving interval in Herefords. *Proc. West. Sec. Am. Soc. Anim. Sci.* 19:307-312.
- Fallon, R. J., F. J. Harte, and D. Harrington. 1987. The effect of calf purchase weight, serum Ig level and feeding systems on the incidence of diarrhea, respiratory disease and mortality. *Bovine Pract.* 22:104-106.
- Goyache, F., J. P. Gutiérrez, I. Alvarez, I. Fernández, L. J. Royo, and E. Gómez. 2003. Genetic analysis of calf survival at different preweaning ages in beef cattle. *Livest. Prod. Sci.* 83:13-20.

- Gregory, K. E., L. V. Cundiff, G. M. Smith, D. B. Laster, and H. A. Fitzhugh. 1978. Characterization of biological types of cattle – Cycle II: I. Birth and weaning traits. *J. Anim. Sci.* 47:1022-1030.
- Kaplan, E. L., and P. Meier. 1958. Nonparametric estimation from incomplete observations. *J. Amer. Stat. Assoc.* 53:457-481.
- Kulkarni, M. D., K. M. Kale, and V. D. Deshmukh. 1994. Effect of birth weight and growth rate on calf mortality in crossbred calves. *Indian Vet. J.* 71:569-571.
- Laster, D. B., H. A. Glimp, L. V. Cundiff, and K. E. Gregory. 1973. Factors affecting dystocia and effects of dystocia on subsequent reproduction in beef cattle. *J. Anim. Sci.* 36:695-705.
- Meijering, A. 1984. Dystocia and stillbirth in cattle. A review of causes, relations and implications. *Livest. Prod. Sci.* 11:143-177.
- Menissier, F., J. L. Foulley, and W. A. Pattie. 1981. The calving ability of the Charolais breed in France, and its possibilities for genetic improvement. 1. The importance and causes of calving difficulties. *Irish Vet. J.* 35:73-81.
- Moore, D. A., W. M. Sischo, D. M. Festa, J. P. Reynolds, E. R. Atwill, and C. A. Holmberg. 2002. Influence of arrival weight, season and calf supplier on survival in Holstein beef calves on a calf ranch in California, USA. *Prev. Vet. Med.* 53:103-115.
- Morris, C. A., G. L. Bennett, R. L. Baker, and A. H. Carter. 1986. Birth weight, dystocia and calf mortality in some New Zealand beef breeding herds. *J. Anim. Sci.*, 62:327-343.
- Pare, J., M. C. Thurmond, I. A. Gardner, and J. P. Picanso. 1993. Effect of birthweight, total protein, serum IgG and packed cell volume on risk of neonatal diarrhea in calves on two California dairies. *Can. J. Vet. Res.* 57:241-246.
- Patterson, D. J., R. A. Bellows, P. J. Burfening, and J. B. Carr. 1987. Occurrence of neonatal and postnatal mortality in range beef cattle. I. Calf loss incidence from birth to weaning, backward and breech presentations and effects of calf loss on subsequent pregnancy rate of dams. *Theriogenology* 28:557-571.
- Ray, D. E., S. B. Itulys, C. B. Roubicek, and C. R. Benson. 1989. Pregnancy rate, calf mortality and calving date in unsupplemented Hereford range cows. *Livest. Prod. Sci.* 23:305-315.
- Rogers, P. L., C. T. Gaskins, K. A. Johnson, and M. D. MacNeil. 2004. Evaluating longevity of composite beef females using survival analysis techniques. *J. Anim. Sci.* 82:860-866.

- Smith, G. M., D. B. Laster, and K. E. Gregory. 1976. Characterization of biological types of cattle. I. Dystocia and preweaning growth. *J. Anim. Sci.*, 43:27- 38.
- Thurmond, M. C. 1986. Epidemiologic approaches used in a herd health practice to investigate neonatal calf mortality. *Prev. Vet. Med.* 4:317-328.
- Yazdi, M. H., P. M. Visscher, V. Ducrocq, and R. Thompson. 2002. Heritability, reliability of genetic evaluations and response to selection in proportional hazards models. *J. Dairy Sci.* 85:1563-1577.

**Appendix: Heritability on the Binary Scale**

The binary data situation is a special case of a discrete measure of survival with only two possible values:  $Y$  is 0 or 1, depending on whether the animal is alive at a given time  $t$  (or, “threshold”). The variable  $Y$  conditional on the breeding value  $s_i$  follows a Bernoulli distribution:

$$Y | s_i \sim Be(p)$$

where the parameter  $p = S(t)$  is the probability that a progeny of sire  $i$  is still alive at time  $t$ . The analysis on the 0/1 scale gives an expectation of  $Y$  conditional on  $s_i$ :

$$E(Y | s_i) = p = S(t)$$

and a variance of  $Y$  conditional on  $s_i$ :

$$V(Y | s_i) = p(1 - p) = S(t)[1 - S(t)]$$

The sire variance  $\sigma_{s(b)}^2$  is obtained using the delta method:

$$\sigma_{s(b)}^2 = V[E(Y | s_i)] = (\text{grad}_{\mu}[S(t)])^2 \sigma_s^2$$

where  $\text{grad}_{\mu}[S(t)]$  is the first derivative  $\frac{d(S(t))}{ds_i}$  evaluated at the mean value  $s_i = \mu_s$ .

Then, the heritability on the binary scale ( $h_{bin}^2$ ) can be approximated as:

$$h_{bin}^2 \approx \frac{4 \sigma_{s(b)}^2}{\sigma_{s(b)}^2 + V_{\mu}(Y | s_i)} = \frac{4 \sigma_s^2}{\sigma_s^2 + \frac{S_{\mu}(t)(1 - S_{\mu}(t))}{(\text{grad}_{\mu}[S(t)])^2}}$$

where the suffix  $\mu$  means that the function is evaluated at the mean breeding value  $s_i = \mu_s$ . Because  $[\text{grad}_{\mu}(S(t))]^2 = [\ln(S_{\mu}(t))]^2 S_{\mu}(t)^2$ , in the case of proportional hazard models, the heritability on the binary scale ( $h_{bin}^2$ ) can be simplified to:

$$h_{bin}^2 \approx \frac{4\sigma_s^2}{\sigma_s^2 + \frac{1 - S_{\mu}(t)}{[\ln(S_{\mu}(t))]^2 S_{\mu}(t)}}$$

This formula is similar to the one derived in a different way by Yazdi et al. (2002) under the assumption of a Weibull distribution. Our derivation only supposes proportionality of hazards and can be applicable to any base distribution.

# **Capítol 4**

## **VALIDATION OF AN APPROXIMATE APPROACH TO COMPUTE GENETIC CORRELATIONS BETWEEN LONGEVITY AND LINEAR TRAITS**

**Joaquim Tarrés<sup>a</sup>, Jesús Piedrafita<sup>a</sup> and Vincent Ducrocq<sup>b</sup>**

<sup>a</sup> *Grup de Recerca en Remugants, Departament de Ciència Animal i dels Aliments,  
Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain.*

<sup>b</sup> *Station de Génétique Quantitative et Appliquée, Institut National de la Recherche  
Agronomique, 78352 Jouy-en-Josas Cedex, France.*

***Genetics Selection Evolution (in press)***

**Abstract**

The estimation of genetic correlations between a nonlinear trait such as longevity and linear traits is computationally difficult on large datasets. A two step approach is proposed and is checked via simulation. First, univariate analyses are performed to get genetic variance estimates and to compute pseudo-records and their associated weights. These pseudo-records are virtual performances free of all environmental effects that can be used in a BLUP animal model, leading to the same breeding values as in the (possibly nonlinear) initial analyses. Combining these pseudo-records in a multiple trait model and fixing the genetic and residual variances to their values computed during the first step, it is possible to obtain correlation estimates by AI-REML and approximate MT-BLUP predicted breeding values that blend direct and indirect information on longevity. Mean genetic correlations and reliabilities obtained on simulated data confirmed the suitability of this approach in a wide range of situations. When nonzero residual correlations exist between traits, a sire model gave nearly unbiased estimates of genetic correlations, while the animal model estimates were biased upwards. Finally, when an incorrect genetic trend was simulated to lead to biased pseudo-records, a joint analysis including a time effect could adequately correct for this bias.

**simulation / genetic correlation / reliability / longevity**



## **1. Introduction**

Functional traits refer to traits related to the ability to remain productive. Their importance increases significantly in situations where production is limited or constrained (quotas) [16]. In general, with the exception of some type traits, functional traits exhibit two problems: rather low heritabilities and insufficient information early in life. These lead to genetic evaluations with low reliabilities for young sires [14]. Fortunately, more heritable traits can be used as early predictors of these functional traits. For example, in dairy cattle, early predictors of, e.g., somatic cells count or functional longevity can be found in the long list of type traits recorded in each breed [2, 6, 18, 21, 22]. A technique to properly combine these pieces of information is needed.

The optimal estimation procedure to combine information from different linear traits is known to be the multiple trait BLUP evaluation [8, 20]. MT-BLUP provides an improved accuracy of the evaluation on each trait through an increase of the amount of information, an improved data structure through better connectedness and a correction of biases due to selection on correlated traits. A multiple trait evaluation automatically accounts for the fact that traits are correlated and that the relative accuracy of the evaluation for each trait may greatly vary between animals [14].

However a MT-BLUP evaluation on functional and production traits altogether, although conceptually possible, is not routinely feasible. Traits are often described by very different models. Some of these models are not linear; others involve repeated measures and/or more than one random effect or are analysed accounting for heterogeneous variances. Above all, the amount of data to manipulate in national evaluations is tremendous. Despite huge and fast improvements in computing power, computational considerations are still a limiting factor. Furthermore, a large set of dispersion parameters must be estimated accurately before being included in such an evaluation.

Ducrocq et al [14] proposed a two-step approach for multiple trait evaluation of functional and production traits. First, univariate analyses are performed for each trait to get genetic variance estimates and to compute pseudo-records and their associated weights. Pseudo-records here can be regarded as a generalization of deviations or records corrected for environmental factors to more complex situations such as repeated records and nonlinear traits. Combining these pseudo-records in a multiple trait animal model while fixing the genetic and residual variances, one can get

correlation estimates and approximate MT-BLUP breeding values that blend direct and indirect information.

Among functional traits, longevity was found to be the most important in most studies on relative economic weights in dairy cattle [5, 16]. Routine genetic evaluations of bulls on length of productive life of their daughters rely on the modelling of a hazard function, which describes the limiting probability for a cow alive just prior to time  $t$  of being culled at time  $t$ . This allows a conceptually natural analysis of records from animals that are still alive (censored records) together with already culled animals (uncensored records). Moreover, this non-linear model used to describe the hazard function can include time-dependent fixed effects (e.g., herd-year-season, stage of lactation) which permit to precisely account for changes in culling policies over time.

The aim of this paper is to check via simulation the two-step approach for multiple trait genetic evaluation of longevity and linear traits. After the analysis of a reference situation, a sensitivity analysis was performed to check the suitability of this approach in a wide range of situations.

## **2. Material and methods**

### **2.1. General strategy**

Observations of longevity ( $t$ ) and of two linear traits ( $y_1, y_2$ ) with known genetic and residual correlations were simulated. For longevity traits, the current models of analysis used are so different (they have to deal with non linearity, censoring and non normal residuals) that an exact multiple trait approach is usually not feasible, at least on moderate or large size data sets. The proposed approach aims at summarising the data in such a way that the simplest linear animal model can be used for each trait. This first step requires the calculation of a pseudo-record  $y_{i,m}^*$  for each animal  $m$  and trait  $i$  corrected for all non genetic effects and an associated weight  $w_{i,m}$  indicating the amount of information for that animal. These pseudo-records are obtained from a univariate (or a simpler multivariate) analysis, after estimation of the relevant dispersion parameters. Then, all pseudo-records are analysed together using a classical MT-BLUP framework assuming an animal model:

$$y_{i,m}^* = \mu_i + a_{i,m} + e_{i,m} \quad (1)$$

where  $\mu_i$  is the overall mean for trait  $i$ ,  $a_{i,m}$  is the additive genetic value of animal  $m$  for trait  $i$  and  $e_{i,m}$  is the residual. To account for the variable amount of information summarised in  $y_{i,m}^*$ , its residual variance is assumed to be heterogeneous:  $\text{var}(e_{i,m}) = \sigma_{e,i}^2 / w_{i,m}$ , where  $\sigma_{e,i}^2$  is the residual variance for trait  $i$ .

The derivation of pseudo-records and their weights is based on the following principle: when analysed using the simplistic univariate BLUP animal model (1), these records should lead to EBVs equal or as close as possible to the EBVs obtained with the complete model and in the case of nonlinear traits, with the adequate methodology. Then, the application of a MT-BLUP animal model based on equation (1) is straightforward. It provides the appropriate EBVs for all traits and all animals. However, MT-BLUP requires an adequate knowledge of the correlations between traits. The variances are supposed to be the ones estimated for the simpler univariate analysis used to compute pseudo-records.

## 2.2. Simulation of the dataset

Records of 5000 animals  $m$  roughly resembling length of productive life of dairy cows were simulated using the following Weibull log normal frailty model:

$$h(t_m) = \lambda \rho (\lambda t_m)^{\rho-1} \exp(\mathbf{x}'_m \boldsymbol{\beta} + \mathbf{z}'_m \mathbf{a}) \quad (2)$$

where  $h(t_m)$  is the hazard function at time  $t_m > 0$ , the Weibull parameters  $\rho$  and  $\lambda$  are strictly positive. A  $\rho$  value of 2.0 (increasing hazard) and a  $\lambda$  value such that the median time  $t_{med}$  was 1000 days were used for the simulation. The two linear traits were simulated based on the model  $y_{i,m} = \mu_i + \mathbf{x}'_{i,m} \boldsymbol{\beta}_i + \mathbf{z}'_{i,m} \mathbf{a}_i + e_{i,m}$ . Two means  $\mu_1 = 100$  and  $\mu_2 = 200$  were also arbitrarily added without lack of generality. For each trait  $i$ , two fixed effects  $\boldsymbol{\beta}'_i = (\boldsymbol{\beta}'_{i,1} \boldsymbol{\beta}'_{i,2})$ , with respectively 10 and 100 unbalanced levels, were generated from a uniform distribution and appropriate change of scale.

The true breeding values of the 5000 animals  $a_{i,m}$ , that were progeny of 50 unrelated sires, were obtained by adding half the breeding value of their sire  $s_{i,m}$  to a value  $u_{i,m}$  covering the dam contribution and mendelian sampling, i.e. representing three quarters of the total genetic variance. Values were drawn from a  $MVN(0, \mathbf{G})$  distribution, where  $\mathbf{G}$  is the genetic covariance matrix. In the reference situation,

genetic variances for longevity and linear characters were respectively 0.20, 400 and 600. Genetic correlations between all pairs of traits were 0.4.

Residual values  $e_{i,m}$  for the two linear traits were also generated from values drawn from a  $BVN(0, \mathbf{R})$  distribution, where  $\mathbf{R}$  is the desired residual covariance matrix. Residual variances were chosen to lead to heritabilities of 0.25 for the first linear character and of 0.10 for the second. The distribution of the residual component for the longevity measure is proportional to an extreme value distribution. The residual correlation between the two linear traits was 0.4. In the reference situation, a residual correlation of 0 was assumed between longevity and the two linear traits.

### 2.3. Calculation of pseudo-records and their associated weight.

Longevity was analysed using a Weibull frailty model [12, 14]. First, the genetic variance was estimated using a sire model. Then, assuming that this estimated variance is the correct one, longevity pseudo-records and their weights were obtained using two procedures. The first one was a two-step procedure where the first step involved the estimation of fixed effects and sire EBVs using a sire model, and the second step consisted in calculating the progeny's EBVs  $\hat{a}_{i,m}$ , considering fixed effects and sire EBVs as known. This results in an approximation of the EBV solutions  $\hat{a}_{i,m}$  from a Weibull animal model (see [9] for details). The second procedure consisted in directly estimating the progeny's EBVs  $\hat{a}_{i,m}$  from a Weibull animal model. The resulting pseudo-record for longevity for animal m in both procedures was (for details, see [15]).

$$y_{i,m}^* = \frac{\delta_m}{w_{i,m}} + \hat{a}_{i,m} - 1 \quad (3)$$

where  $\delta_m = 0/1$  is the censoring code. The associated weight  $w_{i,m} = \hat{H}_{i,m}$  is the cumulative risk  $\hat{H}_{i,m}$  of animal m from time 0 to culling or censoring time. Calculations were done using a modified version (version 5.0) of the Survival Kit [12].

For linear traits, instead of a univariate analysis for each trait, it was decided to use a bivariate animal model, which is simple to implement here. First, genetic and residual dispersion parameters were estimated via an AI-REML procedure. The pseudo-record for trait i and animal m was simply the record corrected for fixed effects:

$$y_{i,m}^* = y_{i,m} - \mathbf{x}'_{i,m} \hat{\boldsymbol{\beta}}_i \quad (4)$$

The associated weight  $w_{i,m}$  for the approximate MT-BLUP evaluation was the diagonal element of the least square part of the mixed model equations (MME) after absorption of the “contemporary group” fixed effect, that is the effect with the largest number of levels. This way, a lower weight is given to observations from small contemporary groups. It may seem inconsistent to absorb this fixed effect on one side and still correct for its estimate in (4). But as pointed out by a referee, the absorption matrix  $\mathbf{M}$  is idempotent ( $\mathbf{M}\mathbf{M}=\mathbf{M}$ ). Therefore, the use of the absorption matrix  $\mathbf{M}$  to weight  $\mathbf{y}^*$  leads to the correct genetic estimates:

$$\begin{aligned} (\mathbf{Z}'\mathbf{M}\mathbf{Z} + \mathbf{A}^{-1}\lambda)\hat{\mathbf{a}} &= \mathbf{Z}'\mathbf{M}\mathbf{y}^* \\ &= \mathbf{Z}'\mathbf{M}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{Z}'\mathbf{M}(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) \\ &= \mathbf{Z}'\mathbf{M}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{Z}'\mathbf{M}\mathbf{M}\mathbf{y} = \mathbf{Z}'\mathbf{M}\mathbf{y} \end{aligned}$$

Using  $\mathbf{y}$  or  $\mathbf{y}^*$  leads to the same solution vector  $\hat{\mathbf{a}}$ .

## 2.4. Joint analysis

Genetic and residual correlations were estimated via an AI-REML procedure applied to pseudo-records assuming that the variances are known and equal to the estimates obtained in the first step (the residual variance for longevity was fixed to 1, since this value plays the role of the residual variance in the variance ratio  $1/\sigma_a^2$  to be multiplied to the inverse relationship matrix [9, equation 12]). Data were analysed using either a sire or an animal model to compare the performance of both models. Several REML packages exist but none is really adapted to model equations such as (1) with heterogeneous residual variances. A simple trick avoids this limitation [14]. Let  $v_{i,m} = \sqrt{w_{i,m}}$ . Multiplying both sides of the model equation (1) by  $v_{i,m}$ , one gets:

$$y_{i,m}^\# = v_{i,m} y_{i,m}^* = v_{i,m}\mu_i + v_{i,m}\mathbf{a}_{i,m} + \varepsilon_{i,m} \quad (5)$$

Now, the residual part  $\varepsilon_{i,m}$  has homogeneous variance:

$\text{Var}[\varepsilon_{i,m}] = v_{i,m}^2 \text{Var}[\mathbf{e}_{i,m}] = \sigma_{e,i}^2$ . The REML estimation of the dispersion parameters of model (5) considering  $y_{i,m}^\#$  as the data and  $v_{i,m}$  as a continuous covariate gives results identical to the analysis of model (1) [14]. A version of I. Misztal's AI-REML software was modified to impose the constraints that genetic and residual variances are fixed [7]. Indeed, this is equivalent to impose structured genetic and residual (co)variance matrices where the only unknown parameters are the correlations. The AI-

REML equations were expressed as functions of the unknown parameters (correlations) and of the first derivatives of the (co)variances matrices with respect to these [7]. Finally, a MT-BLUP evaluation based on pseudo-records (and their weights) was performed using the estimated genetic and residual (co)variance matrices.

## **2.5. Reliabilities**

Two different longevity EBVs were obtained for each animal: one from direct evaluation (via the Weibull model) and the other taking into account indirect information from correlated traits (MT-BLUP approach). Average reliabilities for longevity of progeny and their sires were also computed in two ways. First, asymptotic mean reliability was obtained as the mean of the diagonal elements of the inverse of the Weibull Hessian matrix at convergence of the maximisation process. The second way was to compute the correlation between the simulated (true) BVs and the estimated ones (via Weibull model or via the MT-BLUP approach). The average reliability is the square of this correlation.

## **2.6. Sensitivity analysis**

### **2.6.1. Genetic parameters for longevity**

Two different genetic variances for longevity (either 0.05 or 0.20 (reference)) were used, indicating low and high genetic variation. These variances gave heritabilities for longevity of 0.05 and 0.19 respectively [23].

Also different genetic correlations (0.20, 0.40 and 0.60) were simulated between all three traits (longevity and the two linear traits) as well as a situation where correlations of longevity with linear traits 1 and 2 were equal to 0.4 and -0.4, respectively, and the two linear traits were assumed uncorrelated.

### **2.6.2. Level of random censoring**

To see how censoring affected the results, four levels of random censoring (no censoring and approximately 30, 60 and 90%) were applied to the simulated datasets. Compared with the reference situation (no censoring), random censoring was generated in the following way. All sires from the same dataset had on average the

same percentage of censored records among their progeny. Censoring randomly occurred at time  $C_m$  equal to 400, 800 or 1200 d, mimicking the end of a reproductive cycle. In datasets with a 90% of censored records, all censoring times  $C_m$  were in the 400 days group. In the case of 60% of censored records, 40% of the censoring times  $C_m$  were in the 400 days group, 30% in 800 days and 20% in 1200. For a 30% of censored records, the respective values were 14%, 10% and 7%. Finally, the actual longevity measure available for analysis was set to  $\min(C_m, T_m)$ , where  $T_m$  was the failure time generated as in the reference situation.

### **2.6.3. Batches of progeny with different censoring rates**

The reliability of the younger sires' proofs would increase with the multiple trait approach as a function of censoring percentage. In contrast with the previous section, sires were simulated with different percentages of censoring among their progeny. The actual longevity measure for the first 1000 animals (progeny of the 10 "young" bulls) was set to  $\min(400, T_m)$ , i.e., censored at 400 days if the actual failure time  $T_m$  was higher than 400 days. Longevity for the next 1000 animals was set to  $\min(800, T_m)$  days and the records of the following 1000 animals were set to  $\min(1200, T_m)$  days. Finally, the last 2000 animals were not censored (progeny of old sires).

### **2.6.4. Non zero residual correlation between longevity and linear traits**

So far we assumed a zero residual correlation between longevity and linear traits. Assuming that longevity data follows a Weibull distribution, equation (2) is equivalent to:

$$\log t = \frac{\rho \log \lambda + \mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{a}}{\rho} + \frac{\mathbf{e}}{\rho} \quad (6)$$

where  $\mathbf{e}$  follows an extreme value distribution [17] with  $\text{Var}(\mathbf{e}) = (\pi^2 / 6)$

To generate a nonzero correlation, records for linear traits were simulated as:

$$y_i = \mu_i + \mathbf{x}_i'\boldsymbol{\beta}_i + \mathbf{z}_i'\mathbf{a}_i + \varepsilon_i + \omega_i \quad \text{for} \quad i = 1, 2 \quad (7)$$

where the complete residual values  $e_i$  were decomposed into two components: one ( $\omega_i$ ) correlated with longevity and the other ( $\varepsilon_i$ ) correlated with the other linear trait. That is to say, for  $i = 1, 2$ :

$$\begin{aligned} \text{cov}(\mathbf{e}, \mathbf{e}_i) &= \text{cov}(\mathbf{e}, \omega_i) + \text{cov}(\mathbf{e}, \varepsilon_i) = r_{longi} \sigma_e \sigma_{e_i} + 0 = r_{longi} \sigma_e \sigma_{e_i} \\ \text{cov}(\mathbf{e}_1, \mathbf{e}_2) &= \text{cov}(\varepsilon_1, \varepsilon_2) = r_{12} \sigma_{e_1} \sigma_{e_2} \end{aligned}$$

To obtain  $\omega_{i,m}$ , first the failure time  $y_m$  was simulated in (6). Then, the residual  $e_m$  was obtained as  $e_m = \rho \log t_m + \rho \log \lambda + \mathbf{x}'_m \boldsymbol{\beta} + \mathbf{z}'_m \mathbf{a}$ . The component correlated with longevity  $\omega_{i,m}$  was generated as  $\omega_{i,m} = b_{longi} e_m$ , i.e., the regression of  $e_i$  on  $e$  with  $b_{longi} = r_{longi} (\sigma_{e_i} / \sigma_e)$ .

Finally, the values  $\varepsilon_{i,m}$  were generated from a bivariate normal distribution with the adequate covariance matrix. Note that the non zero residual correlation between longevity and the linear traits was obtained using equation (6) to model  $\log t$  while the final model of analysis with pseudo-records is on a different scale with heterogeneous residual variance. Also, a positive correlation  $r_{longi}$  between linear trait  $i$  and  $\log t$  corresponds to a negative relationship between the linear trait and the hazard. This will have implications on the interpretation of the results.

### **2.6.5. Biases generated by incorrect univariate analyses**

In this section, it is assumed that a genetic trend on all traits existed, i.e. progeny born in years 0, 1, ..., 10 do not have the same average genetic level, and it was incorrectly estimated (biased). To generate such a situation, the simulation models for longevity and linear traits were modified to:

$$h(t_m) = \lambda \rho (\lambda t_m)^{\rho-1} \exp(\mathbf{x}'_m \boldsymbol{\beta} + \delta_j + \mathbf{z}'_m \mathbf{a}) \quad (8)$$

$$y_{i,m} = \mu_{i,m} + \mathbf{x}'_{i,m} \boldsymbol{\beta}_i + \delta_{i,j} + \mathbf{z}'_{i,m} \mathbf{a}_i + e_{i,m} \quad \text{for } i = 1, 2 \quad (9)$$

where, for an animal born in year  $j = 0, 1, \dots, 10$  (11 years), an effect  $\delta_j$  equal to 5% of a genetic standard deviation per year was added for each trait. To create an unbalanced but connected design, each sire was assumed to have half his progeny in year  $j$  and the other half in year  $j + 1$  (where  $j = 0$  to 10). So, sires had their progeny in different years.



The analysis of such a situation was done ignoring the year effect in the first step (univariate analysis) leading to potentially biased estimates of variances, pseudo-records and weights. However, to see how the joint analysis of the data can cope with such biases, a year effect was included in the second step, i.e., for the estimation of the genetic and residual correlations and for the MT-BLUP evaluation of the three traits together.

### **3. Results**

#### **3.1. Reference situation**

Means of the genetic and residual variances estimates over 200 replicates for linear traits were always in the confidence interval of the mean. In the case of longevity, the sire variance is slightly underestimated for the reference situation (0.0449 estimated vs 0.05 simulated). However, this bias may be considered negligible [10].

The calculation of the longevity pseudo-records and their weights was obtained by two alternative procedures: an approximate two-step procedure based on a sire model and a direct procedure based on the exact animal model. The correlations between both procedures were 0.9985 for pseudo-records and 0.9910 for their weights. Therefore, the approximate two-step procedure gave results very similar to the more demanding exact one. Only results from the exact procedure are reported hereafter because the moderate size of our datasets allowed the use of the animal model.

Next, univariate BLUP analyses were performed based on pseudo-records for longevity. The resulting EBVs were compared with the ones calculated from the appropriate Weibull animal model. The correlations between both were 0.9980 for sires and 0.9999 for progeny. Their standard deviations were also nearly identical. As desired, the use of the pseudo-records in the univariate BLUP evaluation based on an animal model lead to sire and progeny EBVs equivalent to the EBVs obtained in the Weibull analysis.

The pseudo-records were useful to compute genetic and residual correlations under a multiple trait sire or an animal model. All mean genetic correlations estimates were similar to the simulated ones whatever the estimation models (Table I). However, it was necessary to impose constraints such that the genetic and residual variances were known and equal to the estimated ones. Otherwise convergence was rarely obtained, due to genetic variances quickly going to 0. The standard deviations of

genetic correlations were between 0.139 and 0.160 for the sire model and were slightly smaller for the animal model (Table I). The average asymptotic standard errors provided by the AI-REML algorithm were lower than these standard deviations (between 0.100 and 0.113 for the sire model). This was expected because it was assumed that the variances were known without error. Nevertheless, these asymptotic standard errors provided a rough idea about the magnitude of the accuracy of the estimates.

**Table I.** Estimates of genetic and residual correlations between longevity (Long) and two linear traits (L1, L2) using the AI-REML approach under a sire and an animal model for two levels of animal genetic variation for longevity: low (0.05) and high (or reference) (0.20).

Model of Analysis		Sire				Animal				
Longevity heritability		Low		High		Low		High		
Correlations <sup>a</sup>		True	Mean <sup>a</sup>	STD <sup>a</sup>	Mean <sup>a</sup>	STD <sup>a</sup>	Mean <sup>a</sup>	STD <sup>a</sup>	Mean <sup>a</sup>	STD <sup>a</sup>
Genetic	Long-L1	0.4	0.382	0.182	0.371	0.139	0.388	0.128	0.374	0.079
	Long-L2	0.4	0.410	0.209	0.389	0.160	0.402	0.179	0.380	0.121
	L1-L2	0.4	0.377	0.158	0.373	0.160	0.379	0.156	0.374	0.156
Residual	Long-L1	0.0	0.004	0.005	0.010	0.006	-0.000	0.006	-0.000	0.006
	Long-L2	0.0	0.003	0.005	0.007	0.006	-0.000	0.005	-0.000	0.006
	L1-L2	0.4	0.406	0.011	0.406	0.011	0.415	0.024	0.416	0.024

<sup>a</sup> Mean and standard deviations (STD) over 200 replicates.

**Table II.** Mean and standard deviation of reliabilities (squared correlations between true and estimated breeding values) of longevity obtained for sires and progeny with a Weibull model and the MT-BLUP approach under a sire and an animal model.

Genetic variance	Model of analysis	Animals	Weibull model		MT-BLUP	
			Mean <sup>a</sup>	STD <sup>a</sup>	Mean <sup>a</sup>	STD <sup>a</sup>
Low (0.05)	Sire	Sires	0.731	0.066	0.743	0.064
		Animal	Sires	0.732	0.066	0.748
	Progeny	0.403	0.043	0.434	0.040	
High (0.20)	Sire	Sires	0.890	0.030	0.892	0.030
	Animal	Sires	0.892	0.030	0.895	0.029
		Progeny	0.538	0.032	0.555	0.029

<sup>a</sup> Mean and standard deviations (STD) over 200 replicates.

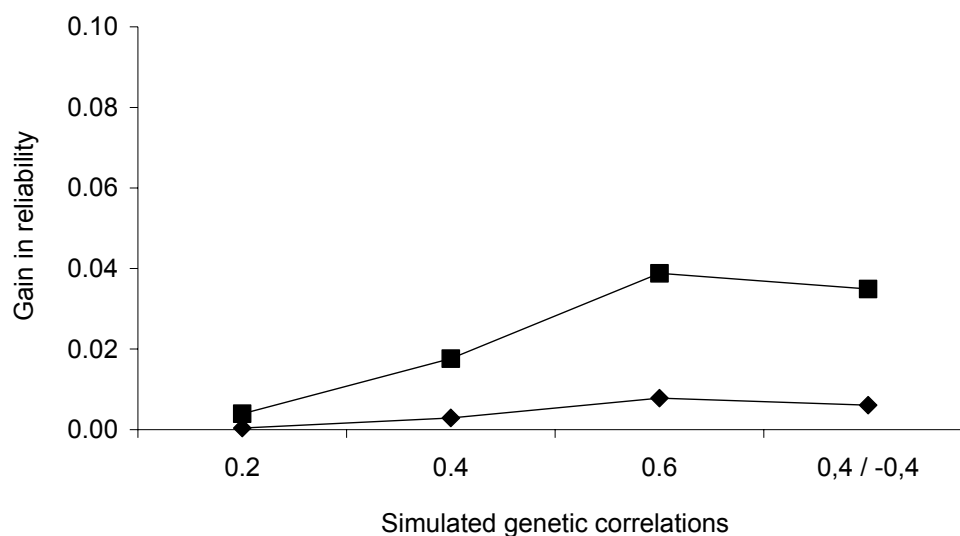
Finally, average reliabilities for longevity computed as the squared correlations between true breeding values and EBVs were 0.890 for sires and 0.538 for progeny when just direct information was used in a Weibull model (Table II). These reliabilities were nearly identical to the asymptotic ones computed from EBV standard errors. When the genetic correlation between longevity and linear characters was accounted for the MT-BLUP approach, the gain in reliability was very limited for sires (0.002 in absolute terms) and slightly higher (0.017) for progeny. This small increase may be related to the initial level of reliability, which was already high for sires.

### 3.2. Sensitivity analysis

#### 3.2.1. Effect of the genetic variance of longevity

Reducing the genetic variance for longevity from 0.20 (reference) to 0.05 did not greatly affect neither the average estimates of genetic variances nor correlations (Table I), although the standard deviations of the latter increased to 0.182-0.209 (sire model). The gain in reliability for the MT-BLUP approach was higher than in the reference situation, 0.012 for sires and 0.031 for progeny (Table II).

**Figure 1.** Increase in reliability for longevity of sires (♦) and progeny (■) obtained by adding indirect information with the MT-BLUP approach under an animal model with respect to the Weibull model. Different levels of genetic correlations are simulated: 0.2, 0.4 and 0.6 and +0.4/-0.4.



### **3.2.2. Effect of the genetic correlation**

Again, the characteristics of the estimates of genetic and residual variances and genetic correlations (results not shown) were not substantially modified by the level of genetic correlation in a range from 0.2 to 0.6. This is also true when genetic correlations of 0.4 between longevity and the first linear trait and -0.4 with the second were imposed. When indirect information was included, the gain in reliability increased with the amount of the genetic correlation simulated (Figure 1). These gains were substantially greater when genetic correlations with different signs were simulated.

### **3.2.3. Effect of level of random censoring**

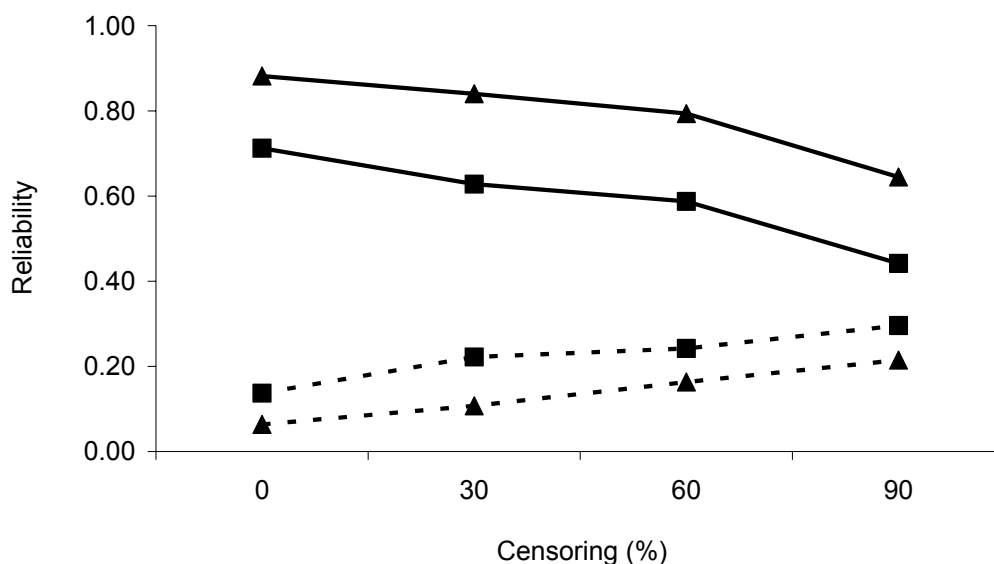
Mean of genetic correlations estimated for datasets with different degrees of censoring were always close to the simulated values. The estimation procedure seems robust even with 90 % of censoring, as the average values were only slightly underestimated. Nevertheless, the amount of censoring clearly affected the standard error of the estimates, which increased moderately until 60% of censoring (from 0.139-0.160 to 0.159-0.196 for the sire model), but the increase was substantial when censoring reached 90 % (0.245-0.294).

Average reliabilities for longevity logically decreased when the level of censoring increased because the amount of direct information decreases. The gains in reliability with the inclusion of indirect information initially increased with censoring rate. In the case of 90 % censoring, there was no gain in reliability for progeny. This is explained by a reduced accuracy of the genetic correlation estimates in this extreme situation as a consequence of the small number of informative records. This was checked by simulating a population of 25000 progeny under the same conditions and with 90% censoring. Then, the standard deviation of the genetic correlation estimates is reduced to 0.08-0.11 (animal model) and the gain in reliability for progeny was about 0.06.

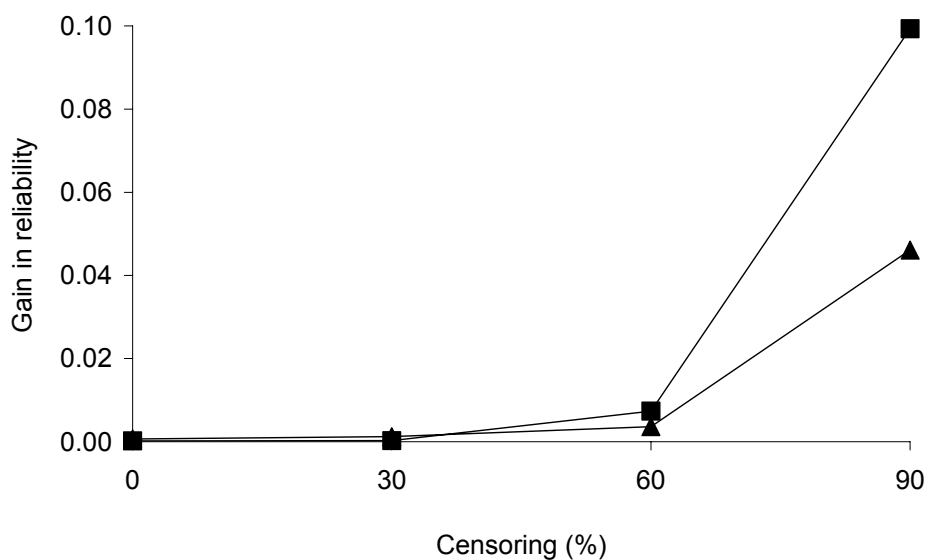
### **3.2.4. Progeny batches with different censoring rate**

When sires were simulated with different percentages of censoring among their progeny, the average reliabilities for longevity decreased and their standard deviations increased when censoring rate increased (Figure 2). The average reliability of young sires with 90 % of their daughters censored was 0.36 and 0.60 for low and high genetic

**Figure 2.** Mean (continuous line) and standard deviations (discontinuous line) of reliabilities for longevity of sires obtained with the Weibull model as a function of censoring percentage among progeny. Different levels of genetic variance for longevity are simulated: low (■) and high (▲).



**Figure 3.** Increase in reliability for longevity of sires obtained by adding indirect information with the MT-BLUP approach as a function of censoring percentage. Different levels of genetic variance for longevity were simulated: low (■) and high (▲).



variation, respectively. When indirect information was taken into account (Figure 3), there was no gain in reliability for old sires (up to 30% censored). However, the gain for young sires was important when genetic variation was high (0.04-0.05), and even more when it was low (up to 0.10). This can be attributed to the fact that the information of older sires allowed a more accurate estimation of variances and genetic correlations compared with the previous situation (all sires with the same percentage of censoring).

### **3.2.5. Effect of non zero residual correlations**

When non zero residual correlations between longevity and linear traits exist, estimates of genetic and residual variances were again similar to the simulated values (results not shown). In the joint analysis, the sire model gave again virtually unbiased estimates of genetic correlations. However, the estimates from the animal model were clearly biased (Table III). The direction of the bias depended on the sign of the residual correlation. The standard deviations were similar to those obtained with zero residual correlation. It should be noted that the estimated residual correlations on longevity trait are not comparable to the simulated ones (Table III). The latter ones corresponded to the modelling of  $\log t$  with a residual variance of  $\pi^2 / 6$  [11]. They are not directly transposable to the pseudo-record scale which assumes a heterogeneous residual variance and corresponds to a modelling of the hazard, not of  $\log t$ . Average reliabilities for sires and progeny were similar to those obtained with zero residual correlation (results not shown).

### **3.2.6. Effect of incorrect univariate analysis**

When the univariate analyses were incorrect (for example, because of the existence of a hidden bias in the estimated genetic trend and/or an incorrect modelling of fixed effects), estimated genetic variances (0.0485 for longevity) were slightly increased compared to the reference situation (0.0449) and residual variances were slightly underestimated. The pseudo-records were also biased. A year effect was included in the joint analysis of the data to try to capture this bias. For longevity, the slope of the regression of the year effects estimated with the MT-BLUP approach on year was 0.0221 (sire model) and 0.0223 (animal model) and was very close to the simulated value (0.0224). Similar results were obtained for the year effects of the two linear traits. With the inclusion of this year effect, and in spite of the use of biased genetic and

residual variances, all genetic correlations were on average similar to the simulated ones for both estimation models (Table III). The standard deviations of these correlations were also similar to those of the reference situation, between 0.141 and 0.171. On the other hand, the average reliabilities for sires and progeny were slightly lower (0.01 to 0.02) than the reference ones.

**Table III.** Estimates of genetic and residual correlations between longevity (Long) and the two linear traits (L1, L2) using the AI-REML approach under a sire and an animal model. Two situations are compared: non zero residual correlation (0.4) and incorrect univariate analysis. A high genetic variation for longevity (0.20) is assumed.

Correlations <sup>a</sup>		Non zero residual			Incorrect univariate		
		True	Sire	Animal	True	Sire	Animal
Genetic	Long-L1	0.4	0.355	-0.371	0.4	0.383	0.359
	Long-L2	0.4	0.360	-0.533	0.4	0.398	0.366
	L1-L2	0.4	0.382	0.554	0.4	0.400	0.394
Residual	Long-L1	0.4 <sup>b</sup>	-0.092 <sup>c</sup>	-0.087 <sup>c</sup>	0	0.010	-0.001
	Long-L2	0.4 <sup>b</sup>	-0.103 <sup>c</sup>	-0.093 <sup>c</sup>	0	0.007	-0.000
	L1-L2	0.4	0.388	0.369	0.4	0.406	0.414

<sup>a</sup> Mean over 200 replicates.

<sup>b,c</sup> Scale change; not comparable to the values on the same line.

<sup>b</sup> corresponds to the correlation between the linear trait and log  $t$ , a negative sign for <sup>c</sup> is expected.

#### **4. Discussion**

Although conceptually possible, the exact joint analysis of longevity data with early predictors or other functional and production traits is not routinely feasible on large data sets. This is due to the need for very different models for these traits (e.g., accounting for nonlinearity, censoring and non normal residuals for longevity traits) and above all, to the fact that the amount of data to manipulate in national evaluations is tremendous. Despite huge and fast improvements in computing power, computational considerations are still a limiting factor. To avoid this, a less demanding two step approach was proposed by Ducrocq [14] and is checked here via simulation. The results obtained under a sire and an animal model confirmed the suitability of the proposed approach in a wide range of situations.

The two step approach starts with the estimation of dispersion parameters via univariate analyses (or simpler multivariate analyses of subsets of traits) and the

evaluation of all recorded animals to get pseudo-records. These pseudo-records are performances free of all environmental effects that can be used in a BLUP animal model to get the same breeding values as in the Weibull animal model. The longevity pseudo-records and their associated weights can be obtained using an animal model if the dataset is small. However, if there are computational constraints to implement it (i.e. for large national applications) a two step procedure to get approximate animal solutions based on a sire model is a less demanding alternative. The correlations between both procedures were very high for pseudo-records as well as for weights and, therefore, both of them could be used indistinctly.

Combining these pseudo-records into a multiple trait sire model and fixing genetic and residual variances to the previously estimated values, AI-REML estimates of genetic and residual correlations were virtually unbiased. This confirms the suitability of the multiple trait approach under a sire model for analysing this kind of data. A concern related to the necessary constraint of assuming that genetic and residual variances are known is that potentially biased genetic and residual variances might lead to biased estimates of correlations. In fact, it was found that the multiple trait approach under a sire model is quite robust: if the genetic trend is wrongly estimated in univariate analyses and estimates of variances, pseudo-records and weights are biased, the joint analysis of the data can correct and even estimate this bias by including a time (year) effect. Again, this leads to nearly unbiased estimates of genetic correlations.

The adequacy of the estimation of genetic correlations in the multiple trait animal model was first assessed under the assumption of a zero residual correlation between longevity and linear traits. This assumption is natural when different traits are recorded in different countries, i.e., on different animals, but is no longer satisfying when the traits are observed on the same animals. Then residual correlations can differ substantially from 0 for some pairs of traits [14]. In such a situation, it was found that the genetic and residual correlations should be estimated under a sire model. The correlation between individual pseudo-record residuals is clearly deviating from the true residual correlation. A sire model somewhat averages residuals over progeny of a same sire and is more robust for variance component estimation. Then, the estimated correlations can be used in a MT-BLUP animal model.

After the MT-BLUP evaluation, a gain in true reliability for longevity is observed with respect to the situation when just direct information is used in a Weibull model. This gain is greater when the reliability in the initial Weibull model is lower. However, this gain was very limited in uncensored datasets, at least with the moderate size of our



dataset. The increase in reliability is more noticeable when progeny batches with different censoring rates are simulated. Then, the information from older sires allows an accurate estimation of genetic variances and correlations and the multiple trait evaluation uses this information to significantly improve the reliability for young sires (with 90% censored daughters).

The suitability of the two step multiple trait approach was also assessed under situations where progeny of all sires had an extreme percentage of censoring. The estimation procedure for variances and correlations seems nearly unbiased but very imprecise in extreme cases, e.g., with 90 % censoring for all progeny groups and small size datasets. In these extreme situations, the gain in reliability is negligible. But the increase in true reliability due to the joint analysis can be quite important, when the dataset is large enough (e.g., at least 25000 records) to accurately estimate variances and correlations. These extreme levels of censoring usually exist only for a fraction of the bulls for dairy cattle length of productive life, but apply more generally to the whole population for example in piglet [4], beef calf [19] or laying hens [13] survival.

It was not our intention to compare this approach with other approximate strategies which, for example, directly combine sets of estimated breeding values using selection index theory. Although computationally more demanding, the strategy proposed here has several attractive features: it accommodates nonlinear traits, it gets as close as possible to a true multiple trait BLUP which has well known theoretical characteristics and it offers a framework to the approximate estimation of genetic correlations between complex traits.

## **Conclusion**

In conclusion, one can note that the two step approach is an operational tool that can be implemented in many situations where a multiple trait approach is desirable but not applicable, either because of the huge size of the datasets analyzed or the complexity and heterogeneity of the models to be implemented. Applications have been reported for total merit index constructions [14], joint analyses of longevity, discrete and linear traits [1], and joint cow and bull international evaluation [3].

## **Acknowledgements**

Some suggestions of the two anonymous referees contributed to significantly improve the manuscript. This work was started during a short stay of J. Tarrés at the Institut National de la Recherche Agronomique in Jouy-en-Josas, France supported by a grant from the “*Ministerio de Educación, Cultura y Deporte*” of Spain.

## **References**

- [1] Besbes B., Ducrocq V., Protais M. An approximate total merit index combining linear traits, a survival trait and a categorical trait in laying hens, in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19-23 August 2002, Montpellier, France, Communication n° 20-05.
- [2] Buenger A., Ducrocq V., Swalve H.H., Analysis of survival in dairy cows with supplementary data on type scores and housing systems from a region of northwest Germany, *J. Dairy Sci.*, 84 (2001) 1531-1541.
- [3] Canavesi F, Boichard D., Ducrocq V., Gengler N., De Jong G., Liu Z., An alternative procedure for international evaluations: production traits european joint evaluation (PROTEJE), in: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, 19-23 August 2002, Montpellier, France, Communication n° 01-59.
- [4] Casellas J, Noguera J.L., Varona L., Sánchez A., Arqué M., Piedrafita J., Viability of Iberian × Meishan F<sub>2</sub> newborn pigs. II. Survival analysis up to weaning, *J. Anim. Sci.* 82 (2004)1925-1930.
- [5] Colleau J.J., Regaldo D., Etablissement de l’objectif de selection dans les races bovines laitières. *Renc. Rech. Ruminants* 8 (2001) 329-332.
- [6] Druet T., Sölkner J., Gengler N., Use of multitrait evaluation procedures to improve reliability of early prediction of survival, *J. Dairy Sci.* 82 (1999) 2054-2068.
- [7] Druet T., Jaffrezic F., Boichard D., Ducrocq V., Modelling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows, *J. Dairy Sci.* 86 (2003) 2480-2490.
- [8] Ducrocq, V., Multiple trait prediction: principles and problems, in: Proceedings of the 5<sup>th</sup> World Congress on Genetics Applied to Livestock Production, 6-11 August 1994, Vol. 18, Guelph, Canada, pp. 452-462.

- [9] Ducrocq V., A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life, *Interbull Bulletin* 27 (2001): 147-152.
- [10] Ducrocq V., An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters, in: *Proceedings of the 55th Annual Meeting of the European Association for Animal Production*, 3-8 September 2004, Bled, Slovenia, Paper G6.11.
- [11] Ducrocq, V., Casella G., A Bayesian analysis of mixed survival models, *Genet. Sel. Evol.* 28 (1996) 505-529.
- [12] Ducrocq, V. P., Sölkner J., The Survival Kit v3.12", a FORTRAN package for large analysis of survival data, in: *Proceedings of the 6<sup>th</sup> World Congress on Genetics Applied to Livestock Production*, 11-16 January 1998, Vol. 27, University of New-England, Armidale, Australia, pp. 447-450.
- [13] Ducrocq V., Besbes B., Protais M., Genetic improvement of laying hens viability using survival analysis, *Genet. Sel. Evol.* 32 (2000) 23-40.
- [14] Ducrocq V., Boichard D., Barbat A., Larroque H., Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index, in: *Proceedings of the 52nd Annual Meeting of the European Association for Animal Production*, 26-29 August 2001, Budapest, Hungary, Paper GI.4.
- [15] Ducrocq V., Delaunay I., Boichard D., Mattalia S., A general approach for international genetic evaluations robust to inconsistencies of genetic trends in national evaluations, *Interbull Bulletin* 30 (2003): 101-111.
- [16] Essl A., Length of productive life in dairy cattle breeding: a review, *Livest. Prod. Sci.* 57 (1998): 79-89.
- [17] Kalbfleisch J.D., Prentice R.L., *The statistical analysis of failure time data*, Wiley, New York, 1980.
- [18] Larroque H., Ducrocq V., Relationships between type and longevity in the Holstein breed, *Genet. Sel. Evol.*, 33 (2001) 39-59.
- [19] Tarres J., Casellas J., Piedrafita J., Genetic and environmental factors influencing mortality up to weaning of Bruna dels Pirineus beef calves in mountain areas. A survival analysis, *J. Anim. Sci.* 83 (2005) 543-551.
- [20] Van der Werf J., van Arendonk J.A.M., De Vries A.G., Improving selection of pigs using correlated characters, in: *Proceedings of the 43rd Annual Meeting of*

the European Association for Animal Production, 14-17 September 1992, Madrid, Spain.

- [21] Visscher P.M., Goddard M.E., Genetic parameters for milk yield, survival, workability and type traits for Australian dairy cattle, *J. Dairy Sci.* 78 (1995) 205-220.
- [22] Weigel K.A., Lawlor T.J., Van Raden P.M., Wiggans G.R., Use of linear type and production data to supplement early predicted transmitting abilities for productive life, *J. Dairy Sci.* 81 (1998) 2040-2044.
- [23] Yazdi, M.H., Visscher P.M., Ducrocq V., Thompson R., Heritability, reliability of genetic evaluations and response to selection in proportional hazards models, *J. Dairy Sci.* 85 (2002) 1563-1577.



## ***DISCUSSIÓ***

## **DISCUSSIÓ**

El programa de millora de la raça *Bruna dels Pirineus* ha de definir i avaluar els caràcters biològics que tenen importància econòmica en la producció. En la primera aproximació a la determinació dels objectius de selecció seguint el model de Phocas i col·l. (1998), els caràcters biològics amb major valor econòmic van ser els següents: interval entre parts, prolificitat, taxa de reforma, supervivència del vedell del naixement fins a l'escorxador, pes al deslletament, creixement durant l'engreix i conformació de la canal (Quintanilla i col·l., 2000). Dins d'aquests, els caràcters reproductius, particularment la prolificitat, juntament amb la taxa de reforma de les vaques i la supervivència dels vedells en diferents períodes eren les variables amb el major pes econòmic en la funció de beneficis (Quintanilla i col·l., 2000). Phocas i col·l. (1998) també afirmaven que els caràcters reproductius eren els caràcters més importants econòmicament en la majoria de sistemes de producció de boví de carn. No obstant, per a expandir i completar el model encara era necessari, entre d'altres, fer un estudi dels diferents factors que afecten la longevitat de les vaques (relacionada amb la taxa de reforma) i la supervivència dels vedells fins al deslletament. Per a fer aquests estudis, en el present treball hem utilitzat les tècniques d'anàlisi de supervivència. En aquesta discussió comentarem els avantatges i inconvenients de l'anàlisi de supervivència i la seva aplicació a l'anàlisi de la longevitat de les vaques, la supervivència dels vedells fins al deslletament i a l'avaluació conjunta de caràcters de supervivència amb caràcters lineals. Finalment exposarem la continuïtat que pot tenir aquest treball en el futur per a l'elaboració d'un índex de selecció mitjançant la derivació dels pesos econòmics tenint en compte l'estructura d'edats i l'estimació conjunta dels paràmetres genètics.

### **Avantatges i inconvenients de l'anàlisi de supervivència**

Històricament, s'han utilitzat diferents tècniques per analitzar les variables que, com la longevitat i la supervivència, mesuren el període de temps fins que es produeix un esdeveniment. Existeixen diferents revisions on s'ha descrit breument aquests procediments d'anàlisi incloent una discussió sobre els avantatges i inconvenients de cada alternativa (Ducrocq, 1987; Strandberg i Sölkner, 1996; Strandberg, 1997; Essl, 1998; Vollema, 1998; Chirinos, 2003). Les principals aproximacions que s'han proposat i aplicat per analitzar aquestes variables són: la mesura binomial de la supervivència,

l'ús de codis de supervivència al llarg de la vida de l'animal, la projecció fenotípica de dades incompletes i l'anàlisi de supervivència.

La primera aproximació consisteix en una modelització de diferents mesures binomials que indiquen si un animal ha superat una determinada edat  $t_0$ . La variable resposta és considerada com un caràcter binari i analitzada amb un model lineal o bé un model llindar (Madgwick i Goddard, 1989; Jairath i col·l., 1998; Vollema i Groen, 1998; Boettcher i col·l., 1999). Els codis de supervivència al llarg de la vida de l'animal són mesures binomials successives de la supervivència dels animals en intervals de temps consecutius, per exemple, anys consecutius. Madgwick i Goddard (1989) proposaren la seva utilització com a mesura per a pal·liar la pèrdua d'informació que suposa l'ús d'una única mesura binomial. Una altra aproximació, la projecció fenotípica de dades incompletes consisteix en fer una modelització lineal combinada amb la predicció dels registres dels animals que es mantenen vius al finalitzar l'estudi. Per a aquesta aproximació, Brotherstone i col·l. (1998) i Van Raden i Klaaskate (1993) desenvoluparen mètodes simples per predir l'esperança de vida del registres censurats a partir de la informació disponible. Llavors aquests registres s'analitzen conjuntament amb els registres complets utilitzant un model lineal.

Per comparar els principals avantatges i inconvenients de l'anàlisi de supervivència respecte als mètodes presentats anteriorment ens basarem en els sis criteris proposats per Veerkamp i col·l. (2001): (i) tractament de dades censurades; (ii) us d'efectes fixes amb efectes diferents al llarg del temps; (iii) reconeixement de la no-normalitat de les dades; (iv) combinació amb altres predictors; (v) nombre d'efectes genètics diferents al llarg del temps i (vi) possibilitat d'ajustar un model animal. Les característiques dels diferents mètodes en relació a aquests sis aspectes les presentarem actualitzant la taula de Chirinos (2003):

Mètode d'anàlisi de la supervivència				
	Mesura binomial	Codis de supervivència	Projecció de dades	Anàlisi de supervivència
i	NO	SI	SI	SI
ii	NO	SI	NO	SI
iii	NO	NO	NO	SI
iv	SI	SI	SI	SI
v	UN	VARIS	UN	VARIS
vi	SI	SI	SI	SI/NO



Les característiques més atractives de l'anàlisi de supervivència són: (i) tota la informació és eficientment usada, tant si el registre és complet com censurat, (ii) ofereix un esquelet convenient per considerar que alguns efectes que influencien el risc varien al llarg del temps de vida, i (iii) permet treballar amb distribucions desconegudes i extremadament asimètriques com sol ser la distribució del temps de vida (Vukasinovic, 1999). Si bé l'ús de codis de supervivència seria una bona alternativa segons la taula, Boettcher i col·l. (1999a) indiquen que l'ús de dades censurades és menys actualitzat que en el cas de les tècniques d'anàlisi de supervivència ja que és necessari esperar fins al següent punt de tall per assignar un nou codi de supervivència, mentre que en l'anàlisi de supervivència la dada és censurada en el moment exacte. La resta de mètodes s'han centrat en solucionar el primer aspecte relacionat amb la utilització de dades censurades i, en canvi, els punts (ii) i (iii) no han estat considerats en aquests procediments. Pel que fa al punt (iii), encara que hi ha hagut treballs on s'ha aplicat un anàlisi no lineal, l'aplicació habitual és sota mètodes lineals aplicables a dades distribuïdes normalment (Chirinos, 2003). Per tant, l'ús de l'anàlisi de supervivència és una alternativa teòricament més adequada per l'anàlisi dels factors que afecten el temps de vida ja que és superior en la modelització de les tres qüestions enumerades.

Les característiques esmentades fan que l'anàlisi de supervivència sigui més apropiat estadísticament i, en conseqüència, el mètode adequat per l'anàlisi de les mesures de longevitat i supervivència en Producció Animal. No obstant, les tècniques d'anàlisi de supervivència no estan limitades a l'anàlisi de caràcters de supervivència i poden aplicar-se a l'anàlisi d'altres tipus de variables contínues o discretes que no segueixen un model lineal normal. De fet, aquestes tècniques ja s'han aplicat a l'anàlisi del temps fins a l'èxit després d'una inseminació (Allore i col·l., 2001), el temps fins la primera mastitis (Carlén i col·l., 2005) o el temps del part fins a l'última inseminació (Schneider i col·l., 2005).

En el camp de la millora genètica, malgrat els avantatges que ofereix en els punts (i), (ii) i (iii), l'anàlisi de supervivència actualment també té desavantatges respecte als altres mètodes quan s'han d'utilitzar models mixtes (models de fragilitat). Un desavantatge era que l'anàlisi de supervivència no permetia combinar el temps de vida amb altres caràcters predictors (Veerkamp i col·l., 2001). No obstant, l'aproximació multicaràcter amb dos passos proposada per Ducrocq i col·l. (2001) permet superar aquesta limitació. En el quart capítol d'aquesta tesi hem validat amb

dades simulades la idoneïtat d'aquesta aproximació en un ampli ventall de situacions (diferents correlacions genètiques, heretabilitats i nivells de censura).

El segon desavantatge segons Veerkamp i col·l. (2001) era que amb l'anàlisi de supervivència no es pot estimar un nombre d'efectes genètics diferents al llarg del temps. L'anàlisi de supervivència assumeix la hipòtesi de proporcionalitat de l'efecte genètic i, per tant, aquest és sempre el mateix al llarg del temps (independent del temps). No obstant, les causes de reforma poden variar clarament dins i fora de cada cicle (per exemple, en vaques lleteres degut a baixa producció lletera, defectes de tipus, fertilitat, malalties) i conseqüentment el component genètic de la reforma també podria variar al llarg del temps (Ducrocq, 2004). La naturalesa multicaràcter d'aquest component genètic queda evidenciada per les baixes estimacions de correlacions entre les supervivències en boví lleter en diferents cicles trobades usant codis de supervivència en els treballs de Jairath i Dekkers (1995) i Veerkamp i col·l., (2001). Si no tenim en compte aquesta interacció amb el temps i mantenim la proporcionalitat de l'efecte genètic com un compromís, el coeficient estimat és com l'efecte mitjà de la variable al llarg dels temps (Allison, 1995). No obstant, si la dependència de l'efecte genètic en el temps fos tan forta que s'hagués d'estudiar, es podria considerar el temps de vida fins a una determinada edat i el temps de vida després d'aquesta edat com a caràcters diferents (per exemple, la supervivència en cada cicle) i després analitzar-los amb una aproximació multicaràcter tenint en compte la correlació entre ells. La validació de l'aproximació multicaràcter amb dos passos torna a ser important en aquest context.

Finalment, un últim desavantatge, i potser el més important actualment, és la impossibilitat pràctica d'ajustar un model animal de supervivència ja que no pot ser utilitzat per a estimar els components de variància, almenys amb l'aproximació de Ducrocq i Casella (1996) (Ducrocq, 2004), a menys que disposem d'una molt bona estructura de pedigrí com, per exemple, en el cas dels conills. Això és degut a la falta d'informació existent en les dades per a realitzar l'estimació de la variància associada a un terme aleatori lligat a l'observació. En canvi, la resta de procediments mencionats no tenen aquest problema ja que s'apliquen en avaluacions genètiques mitjançant la utilització de models lineals i mètodes de tipus BLUP i REML per la valoració genètica i l'estimació dels components de variància (Chirinos, 2003). Degut a la impossibilitat d'aplicar un model animal, el model pare és l'únic model operacional en anàlisi de supervivència per estimar components de variància (Ducrocq, 2004), ja que un model pare fa una mitjana dels residus de la seva progènie i és més robust per a l'estimació

dels components de variància. Amb el software Survival Kit (Ducrocq i Sölkner, 1998) utilitzat en aquest treball, primer s'ha d'estimar la variació genètica amb un model pare i, posteriorment, assumint que la variància estimada és la correcta, es pot procedir a l'avaluació genètica dels animals, ja sigui amb un model pare o amb un model animal. No obstant, encara que el model animal pot ser utilitzat per avaluacions genètiques del temps de vida en poblacions mitjanes, el seu ús planteja problemes computacionals importants quan es manegen fitxers de dades de grans dimensions. Korsgaard i col·l. (1998) proposaren una altra possibilitat com és l'ús de mètodes MCMC dins el marc baiesià en aquests mateixos models, incloent en el seu cas un terme de fragilitat que inclou a més de la variabilitat genètica, un terme residual per ajustar la sobredispersió que produeixen altres factors incontrolats. El problema d'aquesta aproximació és que és lenta i com a conseqüència actualment tampoc es pot aplicar a fitxers de dades de grans dimensions, com són les utilitzades habitualment en l'avaluació genètica de la longevitat.

### ***El problema del model pare***

L'estimació de la variància genètica amb el model pare no sempre és correcta (Daamgard i col·l., 2003). Si les dades segueixen un model pare, com el següent:

$$h(t) = h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'_s \mathbf{s})$$

el risc de reforma depèn d'uns efectes fixes  $\boldsymbol{\beta}$  i un efecte aleatori del pare  $\mathbf{s}$  amb matrius d'incidència  $\mathbf{X}$  i  $\mathbf{Z}_s$ . Llavors, l'estimador de la variància genètica amb un model pare no està esbiaixat perquè el model d'estimació i el model que segueixen les dades és el mateix. L'heretabilitat efectiva en aquest model és (Yazdi i col·l., 2002):

$$h^2 = \frac{4\sigma_s^2}{\sigma_s^2 + 1}$$

on  $\sigma_s^2$  és la variància de l'efecte del pare. En aquests models, les tres quartes parts de la variància additiva que no provenen del pare s'ignoren i estan implícitament englobats dins la variància residual.

En canvi, si les dades segueixen un model animal:

$$h(t) = h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'_a \mathbf{a})$$

ara el risc de reforma depèn dels efectes fixos  $\beta$  i un efecte aleatori additiu de l'animal  $a$  amb matrius d'incidència  $X$  i  $Z_a$ . L'heretabilitat efectiva en aquest model, seguint el plantejament de Yazdi i col·l. (2002), és:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + 1}$$

on  $\sigma_a^2$  és la variància de l'efecte de l'animal. Si la variància genètica es pogués estimar amb un model animal, les estimacions no estarien esbiaixades perquè, novament, el model d'estimació i el model que seguirien les dades seria el mateix. No obstant, quan realitzem l'estimació de la variància genètica amb un model pare, tres quartes parts de la variació genètica no estan incloses en el model i, a més, aquesta part no pot ser inclosa en el residu del model perquè els residus Weibull no segueixen la distribució normal usual (Daamgard i col·l., 2003). Aquesta limitació ha portat a aquests autors a qüestionar la validesa de les estimacions dels paràmetres genètics i dels valors genètics predits amb el model pare. No obstant, malgrat que és cert que l'estimador de la variància del pare,  $\hat{\sigma}_s^2$ , està esbiaixat, aquest problema és més teòric que no pas pràctic. Sortosament, la part de la variància no inclosa en el model és negligible comparada amb les grans variàncies residuals del model de riscos proporcionals (Ducrocq, 2004). Els resultats de diverses simulacions (Meuwissen i col·l., 2002; Ducrocq, 2004) mostren clarament que l'aproximació amb un model pare és molt bona, almenys amb les magnituds usuales de variància genètica de la longevitat. Per tant, el principal inconvenient d'utilitzar un model pare segueix essent més el fet d'ignorar les relacions de parentesc entre femelles que d'ignorar la sobredispersió que genera la part de variabilitat genètica no tinguda en compte en el model (Ducrocq, 1999). No obstant, això no seria cert en els casos en que la variància genètica fóra molt gran. En aquestes situacions poc comunes, s'hauria de corregir el biaix de l'estimador de la variància genètica abans de fer l'avaluació amb el model animal.

Per tal de corregir aquest biaix, els resultats (no publicats) de diverses simulacions ens han permès observar que l'heretabilitat efectiva simulada amb un model animal és aproximadament igual a l'estimada amb un model pare. És a dir:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + 1} \approx \frac{4\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + 1}$$

Per tant, aïllant  $\sigma_a^2$  de l'equació anterior, es pot obtenir la variància genètica correcta  $\sigma_a^2$  a partir de la variància pare esbiaixada  $\hat{\sigma}_s^2$ :

$$\sigma_a^2 = \frac{4\hat{\sigma}_s^2}{1 - 3\hat{\sigma}_s^2}$$

En les nostres simulacions (no publicades) s'ha vist que, encara que té un impacte negligible per variàncies genètiques petites (entre 0,05 i 0,20), aquesta correcció millora substancialment l'estimació de la variància genètica i la predicció dels valors genètics per variàncies superiors al rang anterior (per exemple,  $\sigma_a^2 = 1,20$ ).

Una altra possibilitat per estimar correctament la variància genètica  $\sigma_a^2$  seria evitar el biaix de l'estimació de la variància incloent explícitament en el model les tres quartes parts de la variació genètica que no provenen del pare. El model seria:

$$h(t) = h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}_a' \mathbf{a}) \approx h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{z}_s' \mathbf{s} + \mathbf{z}_u' \mathbf{u})$$

on assumim que l'efecte additiu de l'animal  $\mathbf{a} \sim MVN(0, \mathbf{A}_a \sigma_a^2)$  es divideix en un efecte que prové del pare  $\mathbf{s} \sim MVN(0, \mathbf{A}_s \sigma_s^2)$  i un altre efecte  $\mathbf{u} \sim N(0, 3\sigma_s^2)$  que representa tres quartes parts de la variància genètica i agrupa la influència de la mare i la segregació mendeliana. El problema en aquest punt és que la variància de l'efecte  $u$  s'ha de fixar perquè en cas contrari es tenen problemes de convergència. El que es pot fer és iterar: fixar la variància de l'efecte  $u$  a  $\sigma_u^2 = 3\hat{\sigma}_s^2$ , estimar la variància pare  $\hat{\sigma}_s^2$ , canviar la variància de l'efecte  $u$  a  $\sigma_u^2 = 3\hat{\sigma}_s^2$ , reestimar la variància de pare  $\hat{\sigma}_s^2$ , i així successivament. Després d'unes quantes iteracions, l'estimador de la variància de pare  $\hat{\sigma}_s^2$ , en les nostres simulacions, convergeix a un valor no esbiaixat.

### **Anàlisi de la longevitat de les vaques**

Un cop vistos els avantatges i inconvenients de l'anàlisi de supervivència, la decisió de la seva aplicació depèn del tipus de dades amb què tractem. En el cas de la longevitat de les vaques, com que la utilització de variables dependents del temps és molt important (per exemple, pel canvi de valor de les variables en cada part o per la correcció de l'efecte explotació-any-estació), les característiques de l'anàlisi de supervivència fan que sigui estadísticament més apropiat.

En el primer capítol de la tesi hem vist que la longitud de vida productiva mitjana de les vaques de les explotacions de *Bruna dels Pirineus* és aproximadament de 9 anys i es correspon a una taxa de reposició mitjana de l'11% anual. Aquestes longevitats són similars a les descrites en altres poblacions de boví de carn (Buxadé, 1997; Díaz i col·l., 2002). Aquests alts nivells de longevitat de les vaques suposen uns costos de reposició reduïts i habiliten a la vaca per assolir la seva capacitat màxima de producció quan arriba a la plena maduresa. Els principals factors que redueixen aquesta longevitat són edats al primer part molt llargues, dificultats de part creixents, pesos al naixement dels vedells molt grans, guanys de pes durant la lactació molt petits i intervals entre parts molt llargs. A més a més, hem vist que l'efecte d'aquests factors depèn del valor que prenen en cada part però també de la seqüència de valors en els parts previs.

La inclusió o no d'aquestes variables en el model permet distingir entre longevitat funcional i vertadera per a l'estimació dels valors genètics dels toros. Com hem dit a la introducció, la longevitat funcional és el caràcter que s'hauria de seleccionar ja que permetria evitar la reforma involuntària incrementant el marge de possibilitats per la reforma voluntària. Una manera d'aproximar la longevitat funcional és analitzar les mesures directes de longevitat real amb una correcció de la producció dins del ramat (Ducrocq i col·l., 1988). En boví lleter es sol incloure la producció de llet dins el ramat com a factor de correcció. No obstant, en boví de carn, davant la manca d'una variable que corregeixi la producció per si sola, s'haurien d'incloure totes les variables descrites anteriorment perquè cadascuna contribueix, encara que no completament, a la descripció fenotípica del nivell de producció (Rogers i col·l., 2004). D'aquesta manera, la longevitat de les vaques es podria augmentar indirectament reduint l'edat al primer part, la dificultat de part i l'interval entre parts de les vaques, mantenint els pesos al naixement dels vedells i augmentant el seu guany de pes durant la lactació i directament augmentant la longevitat funcional (o longevitat no deguda a les variables anteriors).

Per poder fer aquesta selecció directa s'haurien d'estimar els paràmetres genètics de longevitat funcional de les vaques. En aquest treball, no ho hem fet perquè el nombre de vaques amb longitud de vida productiva completa i pare conegut era petit per poder estimar la variància genètica amb precisió. No obstant, a la literatura sí que podem trobar estimacions de paràmetres genètics de longevitat usant tècniques d'anàlisi de supervivència en altres poblacions de vaques de carn. Rogers i col·l. (2004), en una població creuada, van estimar variàncies del pare per longevitat

funcional i vertadera de 0,03738 i 0,02813 respectivament. Les estimacions d'heretabilitat corresponents, utilitzant la fórmula de Yazdi i col·l. (2002), eren de 0,14 i 0,11 (Rogers i col·l., 2004). Resultats similars s'han publicat quan la longevitat és tractada com una mesura binomial. En aquest cas, les estimacions d'heretabilitat van de 0,11 a 0,14 en dues explotacions d'Angus avaluades per Snelling i col·l. (1995).

En el futur, per poder estimar aquests paràmetres genètics en la nostra població serà necessari descriure el millor possible la funció de risc base  $h_0(t)$ . En el nostre estudi ha estat estimada no paramètricament en cada explotació com una funció de la longitud de la vida productiva  $t$ . No obstant, validar una distribució paramètrica per a aquesta funció permetria augmentar la precisió de les estimacions dels paràmetres genètics i reduir la demanda computacional. Amb les nostres dades hem intentat validar una distribució de Weibull prenent el  $\ln(-\ln S_{KM}(t))$  en funció de  $\ln t$ , però desafortunadament no s'ajusta a una recta (resultats no mostrats) i, per tant, no seria vàlid assumir una distribució de Weibull en una anàlisi paramètrica. Per tal de superar aquest fet, diversos estudis fets en boví de llet han mostrat l'interès de modelitzar la distribució base no al llarg de tota la vida productiva, sinó dins cada cicle (Ducrocq, 2002; Röxström i col·l., 2003). Aquest interès és degut a que el risc de reforma d'una vaca augmenta amb la vida productiva però també depèn del seu estat fisiològic dins el cicle (o sigui, del temps transcorregut des de l'últim part). Per exemple, suposem dues vaques amb 1000 dies de vida productiva, el risc de reforma de la primera, que fa un mes que ha parit i es troba en plena lactació del vedell, serà inferior que el de la segona, que fa vuit mesos que ha parit i es troba seca i no gestant. En els models utilitzats per aquests autors, la funció de risc base  $h_{0,n}(\tau)$  és diferent per cada cicle  $n$  i depèn del nombre de dies des del part previ ( $\tau = t - t_n$ , on  $t_n$  és el valor de  $t$  en la data de part del cicle  $n$ ).

En el futur s'hauria de comprovar en boví de carn si també augmenta la versemblança quan s'analitza una funció de risc base dins de cada cicle. En aquest cas, per tal de validar una distribució paramètrica, es podria intentar validar una distribució de Weibull en cada cicle. Si això no fóra possible s'hauria de recórrer a funcions paramètriques a trossos en cada cicle. La més simple seria ajustar una funció exponencial a trossos (com la utilitzada en l'anàlisi de la supervivència dels vedells en el tercer capítol) en la qual la funció de risc és diferent entre períodes però constant dins d'aquests. Aquesta funció exponencial a trossos també podria ampliar-se amb un grau de llibertat a una Weibull dependent del temps (també utilitzada en el tercer

capítol). Si això no fóra suficient, es podria assumir una distribució Weibull a trossos que consisteix en ajustar una distribució de Weibull diferent en cadascun dels diferents períodes d'un cicle.

Una altra possible ampliació del model utilitzat per analitzar la longevitat seria l'estudi de les causes de reforma. A la introducció hem vist que la reforma d'un animal es pot produir per diferents causes, bé siguin voluntàries o involuntàries. Al igual que en les truges (Dürr i col·l., 2002; Tarrés i col·l., 2005), les vaques reformades per diferents causes poden estar afectades per diferents factors o per diferents factors en diferents graus. Quan les causes de reforma són d'interès i es volen tenir en compte en l'anàlisi, es pot usar el mètode dels riscos competitius (Kalbfleisch i Prentice, 1980). El concepte general d'aquest mètode és que, quan es produeix la reforma de l'animal per una causa determinada, l'animal deixa d'estar a risc per les altres causes. En aquest cas, les funcions de risc causa-específiques es poden obtenir de la mateixa manera que la funció de risc que no considera la causa però considerant totes les reformes degudes a les causes que no siguin d'interès com a censurades a aquell temps (Allison, 1995). Per tant, partint d'una mateixa base de dades, el percentatge de censura variarà segons la causa de reforma analitzada. D'aquesta manera, podríem estimar la longevitat funcional (o aptitud per evitar la reforma involuntària) d'una manera diferent. Fins ara, per estimar la longevitat funcional incloïem un factor relacionat amb la producció que corregia la reforma voluntària (Ducrocq i col·l., 1988a). En canvi, el mètode de riscos competitius permetria fer una aplicació directa de la definició de longevitat funcional considerant com a censurades les vaques que han estat reformades per causes voluntàries. No obstant, per aplicar aquest mètode de riscos competitius s'hauria d'estimular els ramaders per a que registressin les causes de reforma.

El problema d'aquest mètode de riscos competitius és que s'assumeix que les reformes per les diferents causes no estan correlacionades. Aquest no seria el cas quan una vaca és reformada alhora per dues causes (per exemple, baixa fertilitat i baixa producció). En aquest sentit, una altra aplicació d'una aproximació multicaràcter seria estimar les correlacions genètiques i residuals entre les diferents causes de reforma considerant la reforma per cadascuna d'elles com a caràcters diferents. No obstant, l'aproximació multicaràcter amb dos passos podria no ser adequada en aquest cas ja que, si les causes de reforma estan correlacionades, no es compliria la hipòtesi de censura aleatòria en les anàlisis unicaràcters del primer pas. En aquest cas, s'hauria de recórrer a una aproximació multicaràcter exacta.



### **Anàlisi de la supervivència dels vedells**

En el cas de la supervivència dels vedells fins al deslletament, els avantatges d'utilitzar les tècniques de l'anàlisi de supervivència poden no semblar tan clares com quan s'estudia la longevitat de les vaques. En aquest cas, l'estudi s'aplica únicament al període inicial de la vida i les dades són tan properes a una binomial que una mesura contínua podria semblar que no aporta molta més informació. En aquest sentit, es podria argumentar que una anàlisi d'una mesura binomial de la supervivència en un punt concret podria ser més ràpida, alhora que precisa, fàcil d'entendre i donar més opcions (per exemple, estimar efectes directes i materns amb un model animal). Probablement, aquest motiu ha fet que la majoria dels estudis de supervivència de vedells publicats fins ara (Cundiff i col·l., 1986; Ray i col·l., 1989; Goyache i col·l., 2003) utilitzin aquesta aproximació. No obstant, l'aplicació de les tècniques d'anàlisi de supervivència als vedells lactants és un estudi pioner que permet analitzar el procés de supervivència al llarg de tot el període, del naixement al deslletament, i no únicament en un punt concret. A més, permet modelitzar millor els efectes fixes utilitzant variables dependents del temps per estimar efectes diferents dels factors en diferents períodes de temps. En el camp genètic, existeixen estudis de simulació que demostren que l'avaluació genètica d'un caràcter que mesura un període de temps és més precisa amb les tècniques d'anàlisi de supervivència que amb mesures binomials (Carlén i col·l., 2005). No obstant, és veritat que els altres mètodes més simples permeten l'estimació d'efectes directes i materns sota un model animal (veure Goyache i col·l., 2003), cosa que no és possible amb el programari actual per l'anàlisi de supervivència a menys que assumim que ambdós efectes no estan correlacionats. L'extensió de l'anàlisi de supervivència per a la correcta inclusió d'efectes directes i materns podria ser un repte per al futur pròxim.

### **Anàlisi multicaràcter de la supervivència amb caràcters predictors**

La combinació de dades de longevitat i supervivència amb informació indirecta provenint de caràcters predictors en una aproximació multicaràcter permetria augmentar la precisió de les avaluacions genètiques per als toros joves i els vedells. És per això que en aquesta tesi, hem validat per simulació l'aproximació multicaràcter amb dos passos proposada per Ducrocq i col·l. (2001) en una àmplia varietat de situacions i hem vist que els guanys en precisió per als toros joves i la descendència

són importants aplicant aquesta metodologia. Tot i que la simulació s'ha fet pensant en longevitat de vaques, també pot ser aplicat en el cas de la supervivència dels vedells, ja que, malgrat que en aquest cas tenim un percentatge més gran de censura, també tenim major variància genètica i els fitxers de dades són més grans.

Paral·lelament al nostre estudi, Damgaard i col·l. (2005) han proposat una aproximació baiesiana adoptant un mostreig de Gibbs que també permet inferir les correlacions genètiques i residuals entre un caràcter lineal i un caràcter de supervivència. Tot i que aquesta aproximació és conceptualment exacta, la seva aplicació actualment pot estar limitada per consideracions computacionals. A més, un altre avantatge de l'aproximació multicaràcter amb dos passos és que es va proposar pensant no sols en avaluar conjuntament la longevitat i caràcters lineals sinó qualsevol tipus de caràcter sigui quin sigui el seu model ideal d'anàlisi. Actualment, aquest mètode ja s'aplica en vaques lleteres a França, permetent l'avaluació genètica conjunta de fins a 11 caràcters (Ducrocq i col·l., 2001). També s'ha aplicat en gallines ponedores per combinar caràcters lineals, longevitat i caràcters categòrics (Besbes i col·l., 2002).

Tot i que la validació del mètode multicaràcter ha estat feta simulant una distribució Weibull de les dades, no li manca generalitat, ja que a la introducció hem vist que els pseudo-registres i els seus pesos derivats a partir de qualsevol distribució

base paramètrica es calculen de la mateixa manera,  $y_i^* = \frac{\delta_i}{w_i} - 1 + a_i$  i  $w_i = \hat{H}_i(y_i)$

respectivament. Les diferències en la distribució base només es tenen en compte al calcular la funció de risc acumulat  $\hat{H}_i(y_i)$ . Aquest fet podria fer que aquestes fórmules es poguessin aplicar fins i tot quan desconexem la distribució base de les dades, ja que l'anàlisi amb un model semiparamètric de Cox dóna estimacions molt pròximes a les estimacions obtingudes amb el model paramètric correcte, o sigui,  $a_{i,cox} \approx a_{i,par}$  i

$H_{i,cox}(y_i) \approx H_{i,par}(y_i)$ . Per tant, els pseudo-registres  $y_i^* = \frac{\delta_i}{w_i} - 1 + a_{i,cox}$  i els pesos

$w_i = H_{i,cox}(y_i)$  obtinguts amb el model de Cox podrien ser molt similars als que hauríem obtingut amb el model paramètric correcte desconegut. Aquesta metodologia ja ha estat aplicada per a l'estimació de les correlacions genètiques i residuals entre prolificitat (nombre de nascuts vius i de deslletats) i longevitat funcional en conills (Sánchez i col·l., 2004). En aquest cas, la correlació entre els valors additius estimats amb un model animal de fragilitat semiparamètric i un model animal BLUP unicaràcter

a partir dels pseudo-registres va ser de 0,9988. Aquest resultat sembla que validaria l'aproximació, però potser mancaria una validació amb dades simulades que confirmi la idoneïtat d'aquesta aproximació en un ampli ventall de situacions.

### **Futures aplicacions. Elaboració d'un índex de selecció conjunt**

Com ja hem dit a la introducció, els programes de millora en vacú de carn tenen dos tipus de caràcters a considerar: caràcters de producció (creixement, morfologia, conversió alimentària, qualitat de carcassa) i caràcters funcionals (fertilitat, facilitat de part, caràcter maternal, longevitat). A partir d'aquests caràcters, s'han de definir els objectius de selecció i ponderar-los segons la seva importància econòmica per introduir-los en els diferents índexs de selecció. Per a la confecció dels índexs de selecció és necessari el coneixement dels paràmetres genètics relatius als caràcters d'interès i derivar el seu pes econòmic. Els pesos econòmics serveixen per ponderar els valors genètics estimats de tots els caràcters en l'índex global que maximitza la rendibilitat o productivitat permetent la selecció dels millors animals.

En el futur, seria necessari integrar tots els caràcters d'interès econòmic en la definició d'un objectiu de selecció global, tant els funcionals com els productius. L'anàlisi de supervivència és d'utilitat per introduir la longevitat i la supervivència dels vedells a l'índex de selecció ja que permet fer l'avaluació genètica d'aquests caràcters i podria ajudar a estimar el pes econòmic de tots els caràcters biològics en funció del seu efecte sobre l'estructura d'edats.

### ***Derivació de pesos econòmics: Estructura d'edats***

En el futur immediat, els estudis fets en aquesta tesi seran d'utilitat en el repte d'estimar els pesos econòmics dels caràcters d'interès tenint en compte les variacions que impliquen en l'estructura d'edats. Per a la derivació dels pesos econòmics dels objectius de selecció, una vegada especificat el sistema de producció, s'ha de modelitzar la funció de benefici d'una explotació típica (Ponzoni, 1986; Phocas i col·l., 1998). La funció de benefici s'obté com una funció dels ingressos i les despeses. No obstant, la relació entre els diferents caràcters biològics i les funcions d'ingressos i despeses no és simple. Per exemple, un major guany de pes durant la lactació dels vedells augmenta la quantitat de producte venut que és el vedell deslletat, però també redueix l'eliminació voluntària de les seves mares per baixa producció. Un major pes al

naixement augmenta el pes i la supervivència dels vedells al deslletament, però també augmenta l'eliminació de les vaques, la incidència de parts difícils i l'interval entre parts. La reducció de la incidència de parts difícils, apart de reduir les pèrdues directes causades per distòcia, mortinatalitat, cost de veterinari i requeriment de més mà d'obra i les pèrdues indirectes a llarg termini com la disminució de la producció i problemes de fertilitat i de salut (Luo i col·l., 2002), també permetria millorar indirectament la longevitat de la vaca i la supervivència del vedell durant la lactació. Finalment, anar a buscar una gran regularitat reproductiva (un part per any) permetria augmentar el nombre de vedells nascuts, però també reduiria la reforma voluntària per problemes reproductius. Per tant, en aquesta tesi hem vist que els caràcters d'interès econòmic no solament afecten la quantitat de producte venut sinó també la supervivència en diferents períodes. Per tenir en compte aquest fet al modelitzar la funció de beneficis és necessari establir una relació entre la supervivència  $S(t)$  en diferents períodes i l'estructura d'edats.

En el segon capítol hem començat a treballar en aquest aspecte en el cas de la longevitat de les vaques. El primer benefici econòmic potencial d'augmentar la longitud de la vida productiva de les vaques  $t$  és la reducció de la taxa de reposició,  $TR$ . La relació entre aquestes dues variables és:

$$TR(u) = \frac{1}{E(t)} = \frac{1}{\int_0^u S(t) dt}$$

i depèn del temps màxim d'utilització de les vaques  $u$  que es sol suposar infinit. Cal notar que la taxa de reposició és inversament proporcional a l'esperança de la longitud de la vida productiva  $E(t)$ . Un altre dels beneficis econòmics potencials d'augmentar la longitud de la vida productiva de les vaques és la reducció de la importància relativa de les vaques primíparas: una major supervivència implica un percentatge més gran de vaques que aconseguen la capacitat màxima de producció quan arriben a la plena maduresa. Aquest aspecte ha estat quantificat calculant la mitjana d'una variable (en el nostre cas, el pes al deslletament  $PDM$ ) ponderada per la proporció de vaques en cada interval de vida productiva. Això es fa de la següent manera:

$$PDM = TR(u) \left[ \int_0^u S(t) PD(t) dt \right]$$

i depèn del pes al deslletament dels vedells en funció de la longitud de la vida productiva de la seva mare al moment del part  $PD(t)$ . Així, en el cas de la *Bruna dels Pirineus*, veiem que augmentar el pes al deslletament permetria augmentar la seva

mitjana directament (per exemple, els vedells fills de vaques primíparas seleccionades tindran pesos superiors que els fills de vaques primíparas no seleccionades), però també indirectament augmentant el percentatge de vedells nascuts de vaques en edats més productives. Tot i que en el segon capítol hem vist que aquests canvis són petits per una sola variable, l'edat de la vaca afecta a moltes variables i la suma de petits canvis pot suposar un canvi important.

Per tal de millorar l'aproximació feta en el segon capítol, es pot establir la relació entre la funció de supervivència  $S(t)$  i l'estructura d'edats d'una determinada població. Començarem suposant una població on a cada període anual hi han d'entrar  $n$  animals per compensar els animals reformats i mantenir la seva mida  $N$  constant. Aquest nombre es calcula multiplicant la taxa de reposició  $TR(u)$  per la mida de la població  $N$ :

$$n = N TR(u) = \frac{365,25 N}{\int_0^u S(t) \delta(t)}$$

Aquests  $n$  animals solen ser entrats en diferent nombre cada dia en funció de la distribució de les entrades al llarg de l'any  $f(d)$ :

$$n(d) = n f(d) = N TR(u) f(d)$$

Suposant que la funció de supervivència dels animals  $S(t)$  és la mateixa independentment del dia d'entrada, els animals presents un determinat dia  $d$  seran la suma dels animals que han entrat aquell dia, els animals que sobreviuen dels que van entrar el dia anterior, els animals que sobreviuen dels que van entrar dos dies abans, etc. El nombre d'animals amb temps de vida  $t$  presents en la població un determinat dia  $d$  es calcula com:

$$N_{td}(t, d) = n(d - t) S(t) = N TR(u) f(d - t) S(t)$$

que depèn del nombre d'animals  $n(d - t)$  que havien entrat  $t$  dies abans i la supervivència dels animals  $S(t)$  als  $t$  dies. Aquesta seria la fórmula de l'estructura d'edats conjunta en funció del temps de vida i del dia de l'any  $N_{td}(t, d)$ . Aquesta fórmula, amb una altra nomenclatura ja fa molts anys que s'utilitza en demografia (Lexis, 1875).

Els animals presents a la població generen cada dia una sèrie de costos (també d'ingressos). Aquests costos individuals d'un animal d'una població  $Cl_{td}(t, d)$  poden

variar en funció del temps de vida de l'animal  $t$  i del dia de l'any  $d$ . Per exemple, no menja el mateix una vaca en alta producció de llet 30 dies després del part que una vaca seca 270 dies després del part, ni tampoc val el mateix l'alimentació un dia d'hivern que un dia d'estiu. Per tant, els costos (o els ingressos)  $CP_{td}(t,d)$  generats per tots els animals presents en la població amb temps de vida  $t$  un determinat dia  $d$  es calcularia com:

$$CP_{td}(t,d) = N_{td}(t,d) Cl_{td}(t,d) = N TR(u) f(d-t) S(t) Cl_{td}(t,d)$$

multiplicant el número d'animals  $N_{td}(t,d)$  amb temps de vida  $t$  presents en la població aquell dia  $d$  pels costos individuals  $Cl_{td}(t,d)$  que genera un animal amb temps de vida  $t$  present en la població un determinat dia  $d$ . Aquesta seria la fórmula de la funció de costos conjunta en funció del temps de vida i del dia de l'any.

A partir de la funció de costos (o ingressos) conjunta podem obtenir els costos (o ingressos) anuals de la població com:

$$CP = \int_0^{365,25} \int_0^u CP_{td}(t,d) dt dd = \int_0^{365,25} \int_0^u N TR(u) f(d-t) S(t) Cl_{td}(t,d) dt dd$$

la suma dels costos de cadascun dels animals presents amb temps de vida  $t$  els diferents dies de l'any  $d$ . Per exemple, suposem que volem calcular els costos d'alimentació anuals de les vaques primíparas i sabem que el cost d'alimentació d'una vaca primípara depèn de les necessitats energètiques  $NE(t)$  en funció dels dies transcorreguts després del primer part  $t$  i del preu de la unitat d'energia  $PE(d)$  en funció del dia de l'any  $d$ . Llavors, els costos d'alimentació anuals de les vaques primíparas serien:

$$CA_{v\_prim} = \int_0^{365,25} \int_0^{ip} N TR(u) f(d-t) S(t) NE(t) PE(d) dt dd$$

i, a més de les variables anteriors, dependrien del nombre de vaques  $N$  que té l'explotació, de la taxa de reposició  $TR(u)$ , de la funció de distribució de parts  $f(d)$  i de la supervivència de les vaques primíparas  $S(t)$  i de la durada del primer cicle  $ip$ .

Aquesta manera d'obtenir l'estructura d'edats que hem desenvolupat per una població general, la qual permet el càlcul de la funció de costos i la d'ingressos, es podria aplicar al sistema de producció de la *Bruna dels Pirineus*. En aquest cas, els costos totals de l'explotació serien la suma dels costos de cadascuna de les poblacions que la componen: vaques, vedells en lactació, vedells en engreix i vedelles de reposició (que també es podrien diferenciar en funció del cicle). Això permetria

modelitzar la funció de beneficis i derivar uns pesos econòmics dels diferents caràcters considerant tota la informació disponible.

No obstant, per dur a terme una aproximació d'aquesta envergadura encara ens manca molta informació sobre com evoluciona cadascuna de les variables d'interès econòmic en funció de l'edat de l'animal i del dia de l'any. Primer de tot, per muntar l'estructura d'edats és necessari descriure la funció de distribució de parts en cada cicle. Posteriorment, per modelitzar la funció de beneficis faria falta la següent informació: l'evolució de les necessitats nutritives i la capacitat d'ingestió de les vaques durant el cicle, la producció de llet durant l'alletament, la forma de la corba de lactació i la seva influència en la corba de creixement del vedell fins al deslletament, els costos d'alimentació del vedell des dels quatre mesos fins al deslletament, la corba de creixement durant l'engreix i la ingestió d'aliment (i la seva eficiència) en funció de l'edat, l'evolució dels preus de l'alimentació al llarg de l'any, un coneixement més precís dels preus de la canal dels vedells al llarg de l'any, un coneixement més profund de les característiques càrnies de les vaques reformades, els costos fixes per vaca, etc. En una primera aproximació, alguns d'aquests estudis es podrien treure de la bibliografia (entre d'altres Phocas i col·l., 1998; Villalba i col·l., 2000; Amer i col·l., 2001; Fernández-Perea i Alenda, 2004). Posteriorment, una vegada obtinguda la funció de beneficis, s'haurà de derivar els pesos econòmics que tampoc és una etapa fàcil quan hi ha restriccions com les que existeixen a la Unió Europea.

El problema d'aquesta aproximació podria ser que els càlculs són massa elaborats i detallats per al que es farà amb ells després. Així, al fer aquesta aproximació complexa, s'hauria de comprovar la seva superioritat respecte a altres aproximacions més simples. Una aproximació més simple seria únicament canviar, en el model de Quintanilla i col·l. (2000) i de Phocas i col·l. (1998), la definició de la taxa de reposició per  $TR(u)$  i no considerar la supervivència dels vedells com una variable discreta sinó com una funció de supervivència contínua  $S(t)$  avaluada en el temps d'interès. Aquesta aproximació més simple també permetria tenir en compte la influència dels diferents caràcters biològics sobre la taxa de reposició i la supervivència dels vedells.

### ***Estimació conjunta dels paràmetres genètics***

Una vegada determinats els objectius de selecció, es podria fer una avaluació genètica conjunta d'aquests caràcters, tant els funcionals com els productius, ja que permetria

estimar les correlacions genètiques entre ells i el valor genètic de cada animal amb major precisió. Un cop feta l'avaluació genètica, es podrien integrar els caràcters avaluats a l'índex de selecció. El problema és que no és recomanable integrar a l'índex de selecció els caràcters que posseeixen heretabilitat baixa o poca variabilitat genètica (Alenda, 1994), com són els caràcters de supervivència. En aquest cas, la millora genètica de la longevitat i la supervivència es podria dur a terme per resposta correlacionada a la selecció per altres caràcters. Un estudi interessant seria fer una simulació per avaluar la resposta amb un mètode directe i indirecte alternativament, tant per longevitat com per supervivència, assumint els paràmetres genètics estimats. En aquest sentit, més estudis són necessaris per tal de determinar amb major precisió la seva heretabilitat i la seva correlació amb altres caràcters. Per exemple, un estudi interessant seria detectar quins caràcters morfològics podrien estar correlacionats genèticament amb la longevitat de les vaques i la supervivència dels vedells de raça *Bruna dels Pirineus*.

### **Bibliografia**

- Alenda R. 1994. Objetivo de selección en la raza Avileña-Negra Ibérica. *Bovis* 59,11-26.
- Allison P.D. 1995. *Survival Analysis Using the SAS® System. A Practical Guide*. SAS Institute Inc., Cary, North Carolina.
- Allore H.G., Warnick L.D., Hertl J., Gröhn Y.T. 2001. Censoring in survival analysis: a simulation study of the effect of milk yield on conception. *Prev. Vet. Med.* 49, 223-234.
- Amer P., Simm G., Keane M., Diskin M., Wickham B. 2001. Breeding objectives for beef cattle in Ireland. *Livest. Prod. Sci.* 67, 223-239.
- Besbes B., Ducrocq V., Protais M. 2002. An approximate total merit index combining linear traits, a survival trait and a categorical trait in laying hens. A: Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier (França), Comunicació n° 20-05.
- Boettcher P.J., Jairath L.K., Dekkers J.C.M. 1999. Comparison of methods of genetic evaluation of sires for survival of their daughters in the first three lactations. *J. Dairy Sci.* 82, 1034-1044.



- Brotherstone S., Veerkamp R.F., Hill W.G. 1998. Predicting breeding values for herd life of Holstein-Friesian dairy cattle from lifespan and type. *Anim. Sci.* 67, 405-411.
- Buxadé C. 1997. *Vacuno de carne: aspectos claves*. Mundi-Prensa, Madrid.
- Carlén E., Emanuelson U., Strandberg E. 2005. Survival Analysis for Genetic Evaluation of Mastitis in Dairy Cattle: A Simulation study. A: Proceedings of the 56th EAAP Meeting, Uppsala (Suècia), Commission on Cattle Production, Session CG2.5.
- Chirinos Z.R. 2003. *Desarrollo de un sistema de valoración genética para la longevidad en Ganado Frisón Español*. Tesis doctoral, Universidad Politécnica de Madrid.
- Cundiff L.V., MacNeil M.D., Gregory K.E., Koch R.M. 1986. Between and within-breed genetic analysis of calving traits and survival to weaning in beef cattle. *J. Anim. Sci.* 63, 27-33.
- Damgaard L.H., Korsgaard I.R., Simonsen J., Dalsgaard O., Andersen A.H. 2003. Weibull log-normal frailty models. A: Proceedings of the 54th EAAP Meeting, Roma (Itàlia).
- Damgaard L.H., Korsgaard I.R. 2005. A bivariate quantitative genetic model for a linear Gaussian trait and a survival trait, *J. Anim. Sci.* 83 (Suppl. 1), 25 (Abs.).
- Díaz C., Chirinos Z., Moreno A., Carabaño M.J. 2002. Preliminary analysis of functional longevity in the Avileña Negra Ibérica beef cattle breed. A: Proceedings of the 7<sup>th</sup> World Congress on Genetics Applied to Livestock Production, Montpellier (França), 29, 697-700.
- Ducrocq V. 1987. *An analysis of length of productive life in dairy cattle*. Ph.D. thesis, Cornell University, Ithaca, New York, USA.
- Ducrocq V. 2004. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. A: Proceedings of the 55th EAAP Meeting, Bled (Slovenia), Paper G6.11.
- Ducrocq V. 1999. Two years of experience with the French genetic evaluation of dairy bulls on production-adjusted longevity of their daughters. *Interbull Bulletin* 20, 60-67.
- Ducrocq V., Casella G. 1996. A Bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 28, 505-529.

- Ducrocq V., Sölkner J. 1998. The Survival Kit v3.12", a FORTRAN package for large analysis of survival data. A: Proceedings of the 6<sup>th</sup> World Congress on Genetics Applied to Livestock Production, University of New-England, Armidale (Austràlia), 23, 447-450.
- Ducrocq V., Quaas R.L., Pollak E.J., Casella G. 1988. Length of productive life of dairy cows. 1. Justification of a Weibull model. *J. Dairy Sci.* 71, 3061-3070.
- Ducrocq V., 2002. A piecewise Weibull mixed model for the analysis of length of productive life of dairy cows. A: Proceedings of the 7<sup>th</sup> World Congress on Genetics Applied to Livestock Production, Montpellier (França), Comunicació 20-04.
- Ducrocq V., Boichard D., Barbat A., Larroque H. 2001 Implementation of an approximate multitrait BLUP evaluation to combine production traits and functional traits into a total merit index. A: Proceedings of the 52nd EAAP Meeting, Budapest (Hongria), Paper GI.4.
- Dürr J.W., Schneider M.P., Monardes H.G., Cue R.I. 2002. Competing risks analysis of reasons for disposal in Quebec dairy herds. A: Proceedings of the 7<sup>th</sup> World Congress on Genetics Applied to Livestock Production, Montpellier (França).
- Ensminger M.E. 1977. *Animal Science*. The Interstate Printers and Publishers Co., Danville, IL.
- Essl A. 1998. Length of productive life in dairy cattle breeding: a review. *Livest. Prod. Sci.* 57, 79-89.
- Fagerling P.T., Brinks J.S., Stonaker H.H. 1968. Environmental effects on calving interval in Herefords. A: Proceedings of the Western Section of the American Society of Animal Science 19, 307-312.
- Fernández-Perea M.T., Alenda R. 2004. Economic weights for a selection index in Avileña purebred beef cattle. *Livest. Prod. Sci.* 89, 223-233
- Goyache F., Gutiérrez J.P., Alvarez I., Fernández I., Royo L.J., Gómez E. 2003. Genetic analysis of calf survival at different preweaning ages in beef cattle. *Livest. Prod. Sci.* 83,13-20.
- Jairath L., Dekkers J.C.M. 1995. Operational model for genetic evaluation of functional herd life of Canadian Holsteins. *J. Dairy Sci.* 78 (Suppl. 1): 157 (Abstr.).
- Jairath L., Dekkers J.C.M., Schaeffer L.R., Liu Z., Burnside E.B., Kolstad B. 1998. Genetic evaluation for herd life in Canada. *J. Dairy Sci.* 81, 550-562.

- Kalbfleisch J.D., Prentice R.L. 1980. *The statistical analysis of failure time data*. Wiley, New York.
- Kaplan E.L., Meier P. 1958. Nonparametric estimation from incomplete observations. *J. Amer. Stat. Ass.* 53, 457-481.
- Korsgaard I.R., Madsen P., Jensen J. 1998. Bayesian inference in the semiparametric log normal frailty model using Gibbs sampling. *Genet. Sel. Evol.* 30, 241-256.
- Luo M.F., Boettcher P.J., Schaeffer L.R., Dekkers J.C.M. 2002. Estimation of genetic parameters of calving ease in first and second parities of Canadian Holsteins using Bayesian methods. *Livest. Prod. Sci.* 74, 175-184.
- Madgwick P.A., Goddard M.E. 1989. Genetic and phenotypic parameters of longevity in Australian dairy cattle. *J. Dairy Sci.* 72, 2624-2632.
- Meuwissen T.H.E., Veerkamp R.F., Engel B., Brotherstone S. 2002. Single and multitrait estimates of breeding values for survival using sire and animal models, *Anim. Sci.* 75, 15-24.
- Morris C.A., Bennett G.L., Baker R.L., Carter A.H. 1986. Birth weight, dystocia and calf mortality in some New Zealand beef breeding herds. *J. Anim. Sci.* 62, 327-343.
- Neerhof H.J., Madsen P., Ducrocq V., Vollema A.R., Jensen J., Korsgaard I.R. 2000. Relationships between mastitis and functional longevity in Danish Black and White Dairy cattle estimated using survival analysis. *J. Dairy Sci.* 83, 1064-1071.
- Phocas F., Bloch C., Chapelle P., Bécherel F., Renand G., Ménissier F. 1998. Developing a breeding objective for a French purebred beef cattle selection programme. *Livest. Prod. Sci.* 57, 49-65.
- Ponzoni R.W. 1986. A profit equation for the definition of the breeding objective of Australian merino sheep. *J. Anim. Breed. Genet.* 103, 342-357.
- Quintanilla R., Milán M.J., Díaz C., Goyache F., Piedrafita J. 2000. Selection objectives in three representative Spanish beef cattle breeds. FAIR1 CT95 0702 – Final Report.
- Ray D.E., Itulys S.B., Roubicek C.B., Benson C.R. 1989. Pregnancy rate, calf mortality and calving date in unsupplemented Hereford range cows. *Livest. Prod. Sci.* 23, 305-315.
- Rogers P.L., Gaskins C.T., Johnson K.A., MacNeil M.D. 2004. Evaluating longevity of composite beef females using survival analysis techniques. *J. Anim. Sci.* 82, 860-866.

- Röxström A., Ducrocq V., Strandberg E. 2003. Survival analysis of longevity in dairy cattle on a selection basis. *Genet. Sel. Evol.* 35, 305-318.
- Sánchez J.P., Baselga M., Ducrocq V. 2004. Estimation of the correlation between longevity and litter size. Proceedings of the 8th World Rabbit Congress. Puebla (Mexico).
- Schneider M.P., Strandberg E., Ducrocq V., Roth A. 2005. Survival analysis applied to genetic evaluation for female fertility in dairy cattle. *J. Dairy Sci.* 88, 2253-2259.
- Snelling W.M., Golden R.M., Bourdon R.M. 1995. Within herd genetic analyses of stayability of beef females. *J. Anim. Sci.* 73, 993-1001.
- Strandberg E. 1997. Breeding strategies to improve longevity. A: Proceedings of the 48th EAAP Meeting, Vienna (Austria).
- Strandberg E., Sölkner J. 1997. Breeding for longevity and survival in dairy cattle. *Interbull Bulletin* 12, 111-119.
- Tarrés J., Tibau J., Piedrafita J., Fàbrega E., Reixach J. 2005. Factors affecting longevity in maternal Duroc swine lines. *Livest. Prod. Sci.* (en premsa)
- Van Raden P.M., Klaaskate E.J.H. 1993. Genetic evaluation of length of productive life including predicted longevity of lives cows. *J. Dairy Sci.* 76, 2758-2764.
- Veerkamp R.F., Brotherstone S., Engel B., Meuwissen T.H.E. 2001. Analysis of censored survival data using random regression models. *Anim. Sci.* 72, 1-10.
- Villalba D. 2000. *Construcción y utilización de un modelo estocástico para la simulación de estrategias de manejo invernal en rebaños de vacas nodrizas.* Tesis doctoral, Universitat de Lleida.
- Vollema A.R. 1998. *Selection for longevity in dairy cattle.* PhD. Thesis, Wageningen Agricultural University, Wageningen (Holanda).
- Vollema A.R., Groen A.F. 1998. A comparison of breeding value predictors for longevity using a linear model and survival analysis. *J. Dairy Sci.* 81, 3315-3320.
- Vukasinovic N. 1999. Application of survival analysis in breeding for longevity. *Interbull Bulletin* 21, 3-15.
- Yazdi M.H., Visscher P.M., Ducrocq V., Thompson R. 2002. Heritability, reliability of genetic evaluations and response to selection in proportional hazards models. *J. Dairy Sci.* 85,1563-1577.

## ***CONCLUSIONS***

## **CONCLUSIONS**

D'acord amb els resultats exposats, les principals conclusions extretes d'aquest treball són les següents:

1. La longevitat mitjana de les vaques de carn de la raça *Bruna dels Pirineus* es situa al voltant dels 9 anys de vida productiva que representa un taxa de reposició anual de l'11%. Una disminució d'aquests bons nivells de longevitat provocaria un augment de la taxa de reposició anual i una lleugera disminució de les produccions mitjanes de l'explotació.
2. Aquests nivells de longevitat estan influenciats per diferents factors que afecten el procés de reforma de les vaques. En aquest sentit, s'ha observat que hi ha un major risc de reforma (probablement voluntària) de les vaques mares de vedells amb guany de pes durant la lactació molt petits i amb intervals entre parts molt llargs. Per altra banda, també s'observa un major risc de reforma (probablement involuntària) deguda a edats al primer part molt grans, dificultats de part creixents i pesos al naixement dels vedells molt grans.
3. Els factors que afecten el procés de reforma de les vaques s'han de tractar com a variables dependents del temps ja que el seu valor pot canviar en cada part. No obstant, malgrat que el valor que prenen aquestes variables en cada part té un gran efecte sobre el risc de reforma, el seu efecte també depèn de la seqüència de valors que prenen en els parts previs.
4. L'anàlisi de supervivència és una eina potent per investigar i valorar la supervivència dels vedells de la raça *Bruna dels Pirineus* que és del 96,9% al deslletament. Durant la lactació, la mortalitat diària dels vedells segueix una distribució paramètrica exponencial a trossos que és constant en tres períodes diferents: més important durant els primers quinze dies de vida, menor durant la tercera i quarta setmana i, a partir del mes de vida, la mortalitat és irrellevant.
5. La supervivència dels vedells fins al deslletament també està influenciada per diferents factors, tant ambientals com genètics. Per tal de millorar-la, en el camp ambiental, s'ha de donar una atenció especial als vedells nascuts entre Febrer i Juny i als fills de vaques joves, especialment les primíparas. En el camp genètic, tot i que la variància genètica estimada amb diferents models és alta,

l'heretabilitat trobada en la nostra base de dades és baixa perquè la supervivència estimada va ser molt alta.

6. La millora genètica dels caràcters de supervivència podria beneficiar-se de la resposta correlacionada a la selecció per altres caràcters. En aquest sentit, els resultats de les simulacions han demostrat que l'aproximació multicaràcter amb dos passos pot ser una eina operacional per estimar correlacions genètiques entre caràcters de supervivència i caràcters lineals. La validesa de l'aproximació ha estat comprovada per diferents heretabilitats, correlacions genètiques i residuals, nivells de censura i suposant presència de biaixos en els anàlisis previs.
7. Posteriorment, quan aquestes correlacions genètiques són usades en una avaluació genètica conjunta s'observa una millora significativa de la fiabilitat dels valors genètics predits per al caràcter de supervivència en progenitors joves i en la descendència. Aquest fet augmentaria la resposta a la selecció directa per longevitat i supervivència.