

UNIVERSIDAD POLITÉCNICA DE VALENCIA

DEPARTAMENTO DE CIENCIA ANIMAL



*Departamento de
Ciencia Animal*



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Análisis genético de la longevidad en conejas de producción cárnica.

Constitución y evaluación de una línea Longevo – Productiva de conejos

Genetic analysis of longevity in rabbit does for meat production.

Constitution and evaluation of a Long-lived – Productive rabbit line

TESIS DOCTORAL

Juan Pablo Sánchez Serrano

Valencia, 2005

Director: Manuel Baselga Izquierdo

UNIVERSIDAD POLITÉCNICA DE VALENCIA
DEPARTAMENTO DE CIENCIA ANIMAL

Esta tesis ha sido escrita y presentada como uno de los requisitos para optar al grado de Doctor por el Departamento de Ciencia Animal de la Universidad Politécnica de Valencia.

Valencia, Diciembre de 2005

El Doctorando

El Director

Fdo.: Juan Pablo Sánchez Serrano

Fdo.: Manuel Baselga Izquierdo

A mis padres, a Rafa y Antoñi,
sin vosotros todo hubiese sido imposible.

AGRADECIMIENTOS

Sin duda alguna, el grupo al que primero he de agradecer su colaboración para que este trabajo haya llegado a buen puerto es a mi familia. Durante mi estancia en Valencia, y a pesar de los kilómetros, han estado siempre muy presentes, ofreciéndome su apoyo y sabiduría en los momentos difíciles, y haciendo mejores los buenos. Evidentemente, a ellos no sólo les debo una parte importante de esta tesis, sino también quien soy.

La siguiente persona a la que le debo gran parte de este trabajo es a mi maestro y amigo, Manolo Baselga. Él, además de enseñarme a moverme por el mundo de la mejora genética y la estadística, me ha mostrado, con su propio ejemplo, que el trabajo y el esfuerzo son la base fundamental en la adquisición de nuevos conocimientos. No puedo dejar de mencionar su incondicional apoyo cuando, por asuntos personales, el camino se hizo muy cuesta arriba.

Agustín Blasco es otra persona a quien sin duda debo mucho de este trabajo, él fue quien me introdujo en el campo del análisis estadístico bayesiano. Además, creo que es de especial relevancia, y el momento de agradecer, la labor que realiza con la organización de cursos de genética, programación y estadística bayesiana.

He de manifestar mi gratitud a todos los compañeros que en este momento están o han pasado por la cuarta planta del Departamento. Rosa, Per, Noelia, Migue, Selma, Mariam, Águeda, Toñi, Pilar, Bea, Wagdy, Lorena y Marcial, gracias. Con todos vosotros me une una estrecha amistad. Vuestro carácter contribuye a crear un ambiente de trabajo tan agradable que hace que el esfuerzo del día a día sea mucho más llevadero.

Otro grupo de personas que se merece mi gratitud es el personal de granjas. Sin ellos la parte experimental de este trabajo hubiese sido imposible. Además, son muy buenos los ratos que con ellos he compartido, tanto realizando trabajos de granja como durante los almuerzos.

A Benito Pineda y Adrián Rodríguez les estoy profundamente agradecido por los buenos momentos que compartimos en nuestro “lujoso” piso.

Durante el desarrollo de esta tesis ha habido varios momentos en los que ciertas reuniones han provocado un avance importante y de calidad. Dos de ellas las mantuvimos con el personal de la unidad de cunicultura del IRTA, con el personal de la unidad de cunicultura del INRA y con Vincent Ducrocq. Todos ellos contribuyeron a la

definición de los modelos de análisis, pero además, la fluida correspondencia con Vincent Ducrocq y Miriam Piles ha sido especialmente productiva.

Igualmente he de agradecer el buen trato que recibí durante mi estancia en Foulum (Dinamarca) por parte de todo el personal del Departamento de Genética y Biotecnología del DIAS. De todos ellos he de señalar a las dos personas con las que más estrechamente trabajé, Inge Korsgaard y Lars Damgaard, gracias a ellos muchos aspectos oscuros del análisis de supervivencia rápidamente se esclarecieron.

También agradezco la meticulosidad en la recogida de datos de Manel Forne, y que muy amablemente nos los cediera para llevar a cabo el estudio genético de la longevidad.

A una gran parte del sector cunícola español también le debo que este trabajo se haya podido llevar a cabo. Evidentemente, la lista de cunicultores y técnicos que han contribuido para la fundación de la línea L-P es imposible de detallar con rigor. De todos ellos cabe destacar por su especial apoyo a Chema Pacheco, con quien he hecho bastantes kilómetros en busca de las “conejas longevas”. Otra persona con la que también he compartido muy buenos ratos durante estos viajes es Ceferino Torres, igualmente le estoy muy agradecido. Sinceramente, espero que la contribución que la cunicultura española ha hecho a esta tesis quede en cierta medida compensada con la propia línea fundada.

Finalmente, también es de agradecer las facilidades brindadas por José Solera y Antonio Miguel Pérez al ceder sus instalaciones para el desarrollo de la experiencia de evaluación de la nueva línea.

Valencia 16 de Diciembre de 2005

Juan Pablo Sánchez Serrano

RESUMEN

En primer lugar se revisan las distintas definiciones que para la longevidad se han propuesto y se describen brevemente los métodos de análisis empleados para estimar parámetros genéticos y predecir valores de cría dentro de cada una de estas definiciones. A continuación se presentan en detalle los métodos de análisis de supervivencia, que son los empleados en este trabajo. Finalmente se hace una revisión bibliográfica de las estimas de parámetros genéticos que para longevidad se han obtenido en diferentes especies ganaderas.

Seguidamente se abordan los dos objetivos globales de esta tesis, que son, por un lado, la definición de un modelo para el análisis de la longevidad en conejas de producción cárnica, incluyendo la estimación de parámetros genéticos para este carácter, y por otro la evaluación de una línea de aptitud maternal (L-P), que se ha constituido seleccionando en granjas comerciales hembras que hayan manifestado una longevidad extremadamente alta y una prolificidad al menos media.

Con respecto al primer objetivo se han usado técnicas de análisis de supervivencia para estimar los parámetros genéticos en una subpoblación de la línea V procedente de un núcleo de selección en el que no se llevaba a cabo eliminación por cuestiones productivas. Inicialmente el carácter se definió como el número de días entre el primer intento de cubrición y la muerte/eliminación o censura de la coneja. Con esta definición se observó que el modelo más empleado para modelar datos de longevidad en especies ganaderas (modelo Weibull) no era totalmente adecuado para el caso de conejos de producción cárnica. Para una segunda definición de la longevidad, el número de días entre la primera palpación positiva de una coneja y su muerte/eliminación o censura, el resultado del test para determinar si el modelo de Weibull era adecuado o no también fue negativo. Por lo tanto, para estudiar cómo diferentes factores (entre ellos el genético) influyen en la determinación de la longevidad se decidió emplear modelos de riesgos proporcionales semiparamétricos.

Para la primera definición del carácter, el año estación, definido cada seis meses, mostró un efecto significativo sobre el riesgo de eliminación o muerte. Además, en general, se observó que aquellos niveles de factores propios de hembras que pudieran estar enfermas (número de gazapos nacidos vivos muy bajo) fueron los que tuvieron asociado un mayor riesgo de eliminación o muerte. Con respecto al ritmo reproductivo, se apreció que los retrasos en el primer parto tenían un efecto negativo durante toda la

vida de las hembras. Los retrasos en el segundo parto reducían el riesgo de eliminación o muerte, cuando la comparación entre animales retrasados y no retrasados se hacía en el momento del parto de estos últimos. Este efecto puede explicarse por el elevado riesgo asociado con el momento del parto, especialmente en este ciclo, cuando la coneja aún está en fase de crecimiento. Sin embargo, cuando la comparación se hacía en el momento del parto de las conejas retrasadas, las diferencias desaparecían. Los retrasos en conejas adultas tenían un efecto negativo aunque despreciable. La edad de la coneja en el momento de la primera presentación al macho presentó un efecto significativo, siendo más probable la eliminación o muerte de una hembra cuanto más joven fuese en el momento de la primera cubrición. Se estimó una heredabilidad, en escala del $\log-g(t)$, de 0.053, utilizando un modelo padre-abuelo materno.

Con la segunda definición, el efecto año estación se determinó cada tres meses, e igualmente mostró un efecto significativo, y el ritmo reproductivo de la hembra se estableció más adecuadamente, considerando a lo largo de la vida del animal su estado fisiológico (Lactante, Gestante, Vacía o Gestante-Lactante). Con respecto a los efectos sistemáticos del modelo, en general se llegó a las mismas conclusiones que con la primera definición: los niveles propios de hembras que pudieran estar enfermas (pocos nacidos vivos e infertilidad) son los que tuvieron asociado un mayor riesgo de muerte o eliminación. La heredabilidad estimada en escala de $\log-g(t)$ fue de 0.095, mayor que con la primera definición. Este incremento en la estima de la heredabilidad pudo deberse al mejor modelado de los efectos sistemáticos y al empleo de un modelo animal.

Con respecto a la estimación de las correlaciones genéticas entre la longevidad funcional (usando la segunda definición) y la prolificidad, ni con el número de gazapos nacidos vivos ni con el número de gazapos destetados se obtuvieron estimas significativamente diferentes de cero. La correlación ambiental sólo fue significativa entre la longevidad funcional y el número de gazapos nacidos vivos (-0.11 ± 0.016), estima muy baja aunque favorable.

Seguidamente se describe el proceso de constitución de la nueva línea Longevo – Productiva (L-P), que fue similar al aplicado en la constitución de líneas hiperprolíficas de cerdos y conejos, y además se estudia la potencia estadística del experimento a realizar para evaluarla. La evaluación se llevó a cabo mediante la comparación de su longevidad y sus rendimientos productivos con los de la línea V, línea de reconocida prolificidad y aceptable longevidad. El proceso de creación de la

línea L-P ha sido eficaz, pues ésta mostró una mayor capacidad de supervivencia, especialmente a edades tardías, aunque su prolificidad fue algo menor, especialmente durante los primeros partos. Si la prolificidad de ambas líneas se comparan a lo largo de todos los partos, la línea V tuvo aproximadamente 0.5 gazapos por parto más que la línea L-P. A pesar de las diferencias en longevidad y prolificidad no hubo diferencias en la producción acumulada de gazapos destetados entre ambas líneas. Tampoco se detectaron diferencias en fertilidad.

ABSTRACT

Firstly, the different definitions of longevity and the methods to estimate genetic parameters and to predict breeding values associated with the different definitions are reviewed. Afterwards the Survival Analysis methods are described in detail, since they are used extensively in this thesis. Finally, in the introduction, a bibliographic review has been done on the estimated genetic parameters for longevity in different livestock species.

Afterwards the two main objectives of the thesis are raised, which are on one hand to define a model for the analysis of longevity in rabbit does for meat production, including the estimation of genetic parameters for this trait, and on the other hand the evaluation of a maternal line (L-P), which has been found by selection of outstanding females in longevity and with a prolificacy around the mean of the population from commercial farms.

Regarding the first main objective Survival Analysis techniques have been used to infer genetic parameters in a subpopulation of the V line from a selection nucleus in which no culling based on production performances took place. Initially the trait was defined as the number of days between the first mating and the death/culling or censoring of the female. With this definition it was observed that the model most commonly employed to fit longevity data in livestock species (Weibull model) was not totally adequate for rabbits for meat production. For a second definition of the trait, the number of days between the first positive pregnancy test and the death/culling or censoring date, the result of the test to determine whether the Weibull model fit properly our longevity data was also negative. Thus, in order to study how different factors (including the genetic) determine longevity it was decided to use semiparametric proportional hazard models.

For the first definition of the trait it was observed that the year-season (defined every six months) showed a significant effect over the risk of culling or death. Moreover, in general, it was observed that those levels of factors relating to sick does (very low number of young born alive) were the ones with higher associated risk of culling or death. Regarding the reproductive rhythm, it was observed that does delayed at first parturition had a lower survival probability than non delayed females during their whole lives. A doe delayed at second parturition showed a lower risk of dying/being culled than a non delayed female, when the comparison is done at time of

the parturition of the latter. This is explained by the high risk associated with the parturition, especially at these ages, when the does are still growing. If the comparison is done at the parturition date of the former the differences disappear. Delays at late ages had a negative but negligible effect. The age of the female at first mating showed a significant effect – it was more likely for a young doe to die or to be culled than an older one. An heritability of 0.053, on the log-g(t) scale, was estimated using a sire-maternal grandsire model.

With the second definition of the trait the year-season (defined every three months) also showed a significant effect. The reproductive rhythm of the female it was more properly defined by considering the physiological status of the female during its whole life (Lactating, Pregnant, Empty or Pregnant-Lactating). Regarding the systematic effects in the model, in general, the same conclusion as with the first definition was achieved: Those levels relating to sick does (very low number of young born alive and infertility) were the ones with higher associated risk of culling or death. The estimated heritability, on the log-g(t) scale, was 0.095. This increase in the estimated value for this parameter could be due to a better fitting of systematic effects and because an animal model was used.

Regarding the estimation of the correlations between functional longevity (using the second definition) and prolificacy, it was observed that neither the number of young born alive nor the number of young at weaning showed a genetic correlation statistically significant from zero with functional longevity. The only significant environmental correlation was obtained between functional longevity and the number of young born alive (-0.11 ± 0.016). This estimated value was very low but favourable.

Immediately afterwards, the process of foundation of the new Long-lived – Productive line (L-P) is described, which was similar to the one applied in the foundation of hyper-prolific lines of pigs and rabbits. Moreover, the statistical power achieved in the experiment to evaluate this line is studied. The evaluation was carried out by comparing this line for longevity and for productive performances with the V line, which has a recognized high prolificacy and medium longevity. The process founding the L-P line has been successful, since it showed higher survival ability especially late in life than the V line. Regarding prolificacy traits the line L-P had a lower performance than the V line, especially during the first parturitions. If the prolificacy between both lines was compared during all the parturitions the V line showed 0.5 young more than the L-P. In spite of differences in longevity and prolificacy

no differences in cumulative production of weaned young were observed between the two lines. Differences in fertility were not observed either.

RESUM

En primer lloc, es revisen les distintes definicions que s'han fet del caràcter longevitat i es descriuen breument els mètodes d'anàlisi emprats per a estimar paràmetres genètics i predir valors de cria, dins de cadascuna d'aquestes definicions. Es presenten en detall els mètodes d'anàlisi de supervivència, que són els emprats en aquest treball. Finalment, es fa una revisió bibliogràfica de les estimacions de paràmetres genètics que per a longevitat s'han obtingut en diferents espècies ramaderes.

A continuació, s'aborden els dos objectius globals d'aquesta tesi, que són, d'una banda, la definició d'un model per a l'anàlisi de la longevitat en conilles de producció càrnia, incloent-hi l'estimació de paràmetres genètics per a aquest caràcter i, de l'altra, l'avaluació d'una línia d'aptitud maternal (L-P), que s'ha constituït seleccionant en granges comercials femelles que hagen manifestat una longevitat extremadament alta i una prolificitat almenys mitjana.

Respecte al primer objectiu, s'han usat tècniques d'anàlisi de supervivència per a estimar els paràmetres genètics en una subpoblació de la línia V procedent d'un nucli de selecció en què no es duia a terme eliminació per qüestions productives. Inicialment, el caràcter es va definir com el nombre de dies entre el primer intent de cobriment i la mort/eliminació o censura de la conilla. Amb aquesta definició es va observar que el model més emprat per a modelar dades de longevitat en espècies ramaderes (model Weibull) no era totalment adequat per al cas de conills de producció càrnia. Per a una segona definició de la longevitat, el nombre de dies entre la primera palpació positiva d'una conilla i la mort/eliminació o censura, el resultat del test per a determinar si el model de Weibull era adequat o no també va ser negatiu. Per tant, per a estudiar com diferents factors (entre ells el genètic) influeixen en la determinació de la longevitat es va decidir emprar models de riscos proporcionals semiparamètrics.

Per a la primera definició del caràcter, l'any estació, definit cada sis mesos, va mostrar un efecte significatiu sobre el risc d'eliminació o mort. A més, en general, es va observar que aquells nivells de factors propis de femelles que pogueren estar malaltes (nombre d'errors nascuts vius molt baix) van ser els que van tenir associat un major risc d'eliminació o mort. Respecte al ritme reproductiu, es va apreciar que retards en el primer part van tindre un efecte negatiu durant tota la vida de les femelles. Els retards en el segon part reduïen el risc d'eliminació o mort, quan la comparació entre animals retardats i no retardats es feia en el moment del part de estes últimes. Este efecte pot

explicar-se per l'elevat risc associat amb el moment del part, especialment en este cicle, quan la conilla encara està en fase de creixement. Però, quan la comparació es feia al moment del part de les conilles retardades, les diferències desapareixien. Retards en conilles adultes van tindre un efecte negatiu encara que despreciable. L'edat de la conilla en el moment de la primera presentació al mascle va presentar un efecte significatiu, ja que era més probable l'eliminació o mort d'una femella com més jove fóra en el moment del primer cobriment. Es va estimar una heretabilitat, en escala del $\log-g(t)$, de 0.053, utilitzant un model pare-avi matern.

Amb la segona definició l'efecte any estació es va definir cada tres mesos, i igualment va mostrar un efecte significatiu, i el ritme reproductiu de la femella es va definir més adequadament, considerant al llarg de la vida de l'animal el seu estat fisiològic (lactant, gestant, buida o gestant-lactant). Respecte als efectes sistemàtics del model, en general es va arribar a les mateixes conclusions que amb la primera definició: els nivells propis de femelles que pogueren estar malaltes (pocs nascuts vius i infertilitat) són els que van tenir associat un major risc de mort o eliminació. L'heretabilitat estimada en escala de $\log-g(t)$ va ser de 0.095, major que amb la primera definició. Aquest increment en l'estima de l'heretabilitat va poder deure's al millor modelatge dels efectes sistemàtics i a l'ocupació d'un model animal.

Respecte a l'estimació de les correlacions genètiques entre la longevitat funcional (usant la segona definició) i la prolificitat, ni amb el nombre d'errors nascuts vius ni amb el nombre d'errors deslletats es van obtenir estimacions significativament diferents de zero. La correlació ambiental només va ser significativa entre la longevitat funcional i el nombre d'errors nascuts vius (-0.11 ± 0.016), estimació molt baixa, encara que favorable.

A continuació, es descriu el procés de constitució de la nova línia Longevo – Productiva (L-P), que va ser semblant a l'aplicat en la constitució de línies hiperprolífiques de porcs i conills, i a més s'estudia la potència estadística de l'experiment a realitzar per a avaluar-la. L'avaluació es va dur a terme mitjançant la comparació de la seua longevitat i els seus rendiments productius amb els de la línia V, línia de reconeguda prolificitat i acceptable longevitat. El procés de creació de la línia L-P ha sigut eficaç, perquè va mostrar una major capacitat de supervivència, especialment a edats tardanes, encara que la seua prolificitat va ser quelcom menor, especialment durant els primers parts. Si la prolificitat d'ambdues línies es comparen al llarg de tots els parts, la línia V va tenir aproximadament 0.5 errors per part més que la

línia L-P. A pesar de les diferències en longevitat i prolificitat no hi va haver diferències en la producció acumulada d'errors deslletats entre ambdues línia. No es van detectar tampoc diferències en fertilitat.

TABLA DE CONTENIDOS

CAPÍTULO 1: INTRODUCCIÓN 1

1.1.- DEFINICIONES Y MÉTODOS PROPUESTOS PARA EL ESTUDIO DE LOS DATOS DE LONGEVIDAD 1

 1.1.1.- *Supervivencia hasta un determinado momento* 2

 1.1.2.- *Longitud de la vida productiva* 6

1.2.- ANÁLISIS DE SUPERVIVENCIA 10

 1.2.1.- *Historia de la aplicación del análisis de supervivencia a la mejora genética* 10

 1.2.2.- *Definiciones generales* 14

 1.2.2.1.- *Censura* 14

 1.2.2.2.- *Descripción de los datos de longevidad* 15

 1.2.3.- *Estimación de las funciones que describen los datos de longevidad* 17

 1.2.3.1.- *Métodos paramétricos* 17

 1.2.3.1.1.- *Estimación bajo modelos paramétricos* 17

 1.2.3.2.- *Métodos no paramétricos* 19

 1.2.3.2.1.- *Estimador de Kaplan-Meier de la función de supervivencia y otros estimadores relacionados* 19

 1.2.3.2.2.- *Test de hipótesis no paramétricos* 22

 1.2.4.- *Modelos de regresión* 24

 1.2.4.1.- *Extensión de los modelos de riesgos proporcionales* 27

 1.2.4.1.1.- *Estratificación* 27

 1.2.4.1.2.- *Covariables dependientes del tiempo* 27

 1.2.4.2.- *Modelos de riesgos proporcionales paramétricos* 28

 1.2.4.3.- *Modelos de riesgos proporcionales semiparamétricos* 28

 1.2.4.3.1.- *Modelo de Cox* 29

 1.2.4.3.2.- *Piecewise constant hazard model* 31

 1.2.5.- *Modelos de fragilidad* 32

 1.2.5.1.- *Estimación bajo modelos de fragilidad* 35

 1.2.5.2.- *¿Es necesario incluir un residuo en el término de fragilidad?* 37

 1.2.5.3.- *Definiciones de heredabilidad* 39

 1.2.5.3.1.- *Heredabilidad en la escala de log-frailty* 39

 1.2.5.3.2.- *Heredabilidad en la escala de log-g(t)* 40

 1.2.5.3.3.- *Heredabilidad en la escala original* 41

 1.2.5.3.4.- *Heredabilidad efectiva* 43

 1.2.6.- *Análisis conjunto de caracteres de supervivencia y caracteres gaussianos* 45

 1.2.6.1.- *Transformación de los datos de longevidad para poder explicarlos a través de las ecuaciones del modelo mixto* 47

 1.2.6.2.- *Métodos que incluyen un residuo en el término de fragilidad del modelo de supervivencia* 48

1.3.- REVISIÓN BIBLIOGRÁFICA DE LA LONGEVIDAD POR ESPECIES 50

 1.3.1.- *Vacuno Lechero* 50

 1.3.1.1.- *Importancia económica* 50

 1.3.1.2.- *Longevidad verdadera y funcional* 51

 1.3.1.3.- *Parámetros genéticos y selección* 52

 1.3.2.- *Vacuno de carne* 54

 1.3.2.1.- *Importancia económica* 54

 1.3.2.2.- *Variabilidad, heterosis y parámetros genéticos* 55

 1.3.3.- *Pequeños rumiantes* 57

 1.3.4.- *Cerdos* 58

 1.3.4.1.- *Importancia económica* 58

 1.3.4.2.- *Variabilidad genética y selección* 59

 1.3.5.- *Especies de laboratorio* 62

 1.3.6.- *Conejo de carne* 63

 1.3.6.1.- *Causas de eliminación y estructura de la población* 63

 1.3.6.2.- *Importancia económica de la longevidad o de la tasa de reposición* 65

 1.3.6.3.- *Variabilidad, parámetros genéticos y relación con la producción* 66

1.4.- CONCLUSIONES 70

1.5.- BIBLIOGRAFÍA 72

CHAPTER 2: OBJECTIVES 89

CHAPTER 3: GENETIC ANALYSIS OF LONGEVITY IN RABBITS..... 93

 3.1.- *ANALYSIS OF FACTORS INFLUENCING LONGEVITY OF RABBIT DOES* 93

3.1.1.- <i>Abstract</i>	93
3.1.2.- <i>Introduction</i>	93
3.1.3.- <i>Material and Methods</i>	95
3.1.4.- <i>Results and Discussion</i>	98
3.1.4.1.- <i>Checking the proportional hazards assumption</i>	98
3.1.4.2.- <i>Checking the suitability of a Weibull baseline hazard function</i>	99
3.1.4.3.- <i>Defining the model</i>	99
3.1.4.4.- <i>Variance components estimation</i>	100
3.1.4.5.- <i>Effects estimation</i>	101
3.1.5.- <i>Conclusions and implications</i>	104
3.1.6.- <i>Acknowledgements</i>	105
3.1.7.- <i>References</i>	105
3.2.- ANALYSIS OF RABBIT DOES LONGEVITY USING A SEMIPARAMETRIC LOG-NORMAL ANIMAL FRAILTY MODEL WITH TIME-DEPENDENT COVARIATES	109
3.2.1.- <i>Abstract</i>	109
3.2.2.- <i>Introduction</i>	109
3.2.3.- <i>Material and Methods</i>	111
3.2.3.1.- <i>Data</i>	111
3.2.3.2.- <i>Methods</i>	112
3.2.3.2.1.- <i>Model</i>	112
3.2.3.2.2.- <i>Conditional Likelihood</i>	114
3.2.3.2.3.- <i>Prior specification</i>	114
3.2.3.2.4.- <i>Fully conditional posterior distributions and implementation of the Gibbs Sampler</i> ..	114
3.2.4.- <i>Results</i>	117
3.2.4.1.- <i>Variances components</i>	117
3.2.4.2.- <i>Regresion parameters</i>	118
3.2.5.- <i>Discussion</i>	121
3.2.6.- <i>Acknowledgements</i>	124
3.2.7.- <i>References</i>	124
3.3.- GENETIC AND ENVIRONMENTAL CORRELATIONS BETWEEN LONGEVITY AND LITTER SIZE IN RABBITS	127
3.3.1.- <i>Abstract</i>	127
3.3.2.- <i>Introduction</i>	127
3.3.3.- <i>Material and Methods</i>	128
3.3.3.1.- <i>Animals</i>	128
3.3.3.2.- <i>Statistical Methods</i>	129
3.3.3.3.- <i>Calculation of “pseudo-records” and their associated weight</i>	130
3.3.4.- <i>Results and Discussion</i>	132
3.3.4.1.- <i>Heritabilities and correlations for litter size traits</i>	132
3.3.4.2.- <i>Heritability of longevity</i>	133
3.3.4.3.- <i>Validation of the “pseudo-records”</i>	133
3.3.4.4.- <i>Estimation of the correlations between litter size traits and longevity</i>	133
3.3.5.- <i>Acknowledgements</i>	135
3.3.6.- <i>References</i>	135
CHAPTER 4: CONSTITUTION AND EVALUATION OF THE L-P LINE	141
4.1.- CONSTITUTION OF THE L-P LINE	141
4.2.- EXPERIMENTAL DESIGN	144
4.3.- EVALUATION OF THE L-P LINE	149
4.3.1.- <i>Abstract</i>	149
4.3.2.- <i>Introduction</i>	149
4.3.3.- <i>Material and Methods</i>	151
4.3.3.1.- <i>Animal Material and Management</i>	151
4.3.3.2.- <i>Traits Definition and Statistical Models</i>	152
4.3.3.2.1.- <i>Functional Longevity</i>	152
4.3.3.2.2.- <i>Prolificacy Traits</i>	155
4.3.3.2.3.- <i>Cumulative Production</i>	156
4.3.3.2.4.- <i>Interval between positive pregnancy tests</i>	157
4.3.4.- <i>Results</i>	158
4.3.4.1.- <i>Functional Longevity</i>	158
4.3.4.2.- <i>Prolificacy traits</i>	162
4.3.4.3.- <i>Cumulative Production</i>	163
4.3.4.4.- <i>Interval between positive pregnancy tests</i>	164

4.3.5.- <i>Discussion and Conclusions</i>	165
4.4.- REFERENCES	168
CHAPTER 5: DISCUSSION AND CONCLUSIONS.....	173
5.1.- REFERENCES	176

ÍNDICE DE FIGURAS Y TABLAS

Figure 3.1.4.1.1.- Graphical test of the proportional hazard assumption between different generations. SKM_n(t), estimator of Kaplan-Meier of the survival function for generation n; time is days since first presentation to a male..... 98

Figure 3.1.4.2.1.- Graphical test of the assumption of a Weibull distribution for the baseline hazard function. SKM (t), estimator of Kaplan-Meier of the survival function for the whole population; time is days since first presentation to a male..... 99

Figure 3.1.4.5.1.- RR, relative risks for the year-season (YS) effect. 102

Figure 3.1.4.5.2.- Survival function of does with different reproductive patterns. 104

Figure 3.2.4.1.1.- Marginal posterior distributions for h_{nor}^2 and h_{log}^2 118

Figure 3.2.4.2.1.-Posterior means of risks associated with the different levels of PS*BA. 120

Figure 3.2.4.3.1.- Log posterior mean of $H_0(t)$ (Baseline cumulative hazard) versus $\ln(t)$ (days). 121

Figure 4.1.1.- Constitution of the L-P (Long lived - Productive) line. In each step appears the number recruited of L-P does and the total number of male offspring obtained (both sexes in generation 0). In brackets the number of L-P females, actual dams of the offspring indicated. 143

Figure 4.2.1.- Survival functions for the reference population (line V) ($S(t)_{KM}$), and the two populations with different hazard ratios ($[S(t)_{KM}]^{\gamma_{R1}}$) and ($[S(t)_{KM}]^{\gamma_{R2}}$)..... 146

Tabla 1.4.1.- Resumen de la estimas de heredabilidades en las distintas especies ganaderas 70

Table 3.1.3.1.- Minimum, maximum and average productive life for censored and uncensored records in the data set. 95

Table 3.1.4.3.1.- Likelihood ratio test, including factors sequentially added to the fixed effects model..... 100

Table 3.1.4.5.1.- Estimates within reproductive rhythm (RR) of the hazard coefficient (Coef.), standard error (S.E.), relative risk (rr) and number of uncensored data (n_{unc}) for each level of born alive (BA)..... 102

Table 3.2.4.1.1.- Posterior mean (PM), posterior standard deviation (PSD), smallest interval with the 95 % of the density (95% highest posterior density region) (95-SI), effective sample size (ESS) and lag 1 and lag 10 autocorrelations for the chains of the variance components σ_a^2 and σ_e^2 , and the heritabilities h_{nor}^2 and h_{log}^2 117

Table 3.2.4.2.1.- Posterior means (posterior standard deviations) of the risks associated with the different levels of OPPT*PS 119

Table 3.2.4.2.2.- Posterior standard deviations of the risks associated with the different levels of PS*BA	120
Table 3.3.4.1.1.- Estimate and standard error of the heritability (h^2) and of the ratio between the variance of non-additive plus permanent environmental effects and the phenotypic variance (p^2) for the two litter size traits.....	132
Table 3.3.4.4.1.- Variance components used to estimate the correlations for the litter size traits and longevity	134
Table 3.3.4.4.2.- Genetic and environmental correlations between longevity and litter size traits (asymptotic standard error assuming that the variances used are correct).	134
Table 4.2.1.- Number of animals required to detect as significant two different log-hazard ratios for different Type I and Type II errors.....	146
Table 4.2.2.- Simulation check of the statistical power of the different experimental sizes for the two log-hazard ratios assumed	147
Table 4.3.3.1.1.- Summary statistics for the number of parturitions and for the mean number of young born alive of the females which contributed to the foundation of the L-P line.	151
Table 4.3.4.1.1.- Count and Proportion of censored/uncensored data by genetic type (L-P and V).....	158
Table 4.3.4.1.2.- Summary statistics for longevity (in days) of censored/uncensored data by genetic type (L-P and V).	159
Table 4.3.4.1.3.- Summary of number and proportion of females alive at the end of each cycle by genetic type (L-P and V) at the end of the experiment.	159
Table 4.3.4.1.4.- Proportion of females culled or death within the major culling/dead causes.....	160
Table 4.3.4.1.5.-Marginal posterior statistics for the log-hazard ratios between genetic types effects (L-P and V) within cycle order (negative values indicate greater risk associated with line V, implying lower longevity).....	160
Table 4.3.4.1.6.- Marginal posterior statistics (in days) of the mean survival time restricted to 330 days, by genetic type (L-P and V) and for the contrast between both genetic types, for three levels (low, medium and high) of hazard.....	161
Table 4.3.4.2.1.- Summary statistics of the prolificacy traits within genetic type (L-P and V).	162
Table 4.3.4.2.2.- Contrast between genetic types(L-P and V) within cycle for prolificacy traits.	163
Table 4.3.4.2.3.- Contrast between genetic types (L-P and V) for prolificacy traits....	163

Table 4.3.4.3.1.- Summary statistics of cumulative production (number of young at weaning) for censored and uncensored data within genetic type (L-P and V)..... 164

Table 4.3.4.4.1.- Summary statistics of the interval (in days) between positive pregnancy tests for censored and uncensored data within genetic type (L-P and V).
..... 164

Table 4.3.4.4.2.- Marginal posterior statistics for the log-hazard ratios between genetic types (L-P and V) within cycle order (negative values indicate greater risk to get a positive pregnancy test associated with V line, implying higher fertility)..... 165

CAPÍTULO 1: INTRODUCCIÓN

CAPÍTULO 1: INTRODUCCIÓN

De manera general, la permanencia en el rebaño de los animales destinados a la producción es un carácter de interés en ganadería. Si se consigue aumentar, los costes directos de reposición se pueden dividir entre un mayor número de unidades de producto obtenidas por animal presente en la explotación, y además se evitan una serie de problemas productivos en los que se incurre por el mero hecho de hacer la reposición. Como consecuencia de la importancia que la longevidad tiene en producción animal, diferente en cada especie, existe una amplia bibliografía que trata este carácter en la mayoría de las especies ganaderas. Con respecto al conejo de producción cárnica, en particular, los trabajos en los que se aborda el estudio de la longevidad de las reproductoras son muy escasos, a pesar de que en esta especie la tasa de reposición anual es muy elevada y pudiera ser muy interesante determinar los factores, tanto ambientales como genéticos, que determinan la longevidad de las hembras. De esta manera se podrían establecer pautas de manejo (alimenticio, sanitario, etc.) o programas de mejora genética, para tratar de reducir esta elevada tasa de reposición. En esta tesis nos centraremos en la segunda alternativa.

La longevidad es un carácter complejo cuyo estudio se ha abordado siguiendo diferentes definiciones del carácter y metodologías estadísticas, muchas de ellas desarrolladas específicamente para afrontar su estudio. Por lo tanto, antes de abordar una revisión de la importancia del carácter y de las estimas de parámetros genéticos en diferentes especies ganaderas, se describirán las diferentes definiciones y métodos estadísticos propuestos para estudiarlo, desarrollándose con cierto detalle los métodos de análisis de supervivencia, que son los que constituyen el marco metodológico de esta tesis.

1.1.- Definiciones y métodos propuestos para el estudio de los datos de longevidad

En general, la metodología empleada para el análisis de los datos de longevidad va a ir muy ligada a la definición del carácter. De entrada esta definición debiera de hacerse de tal manera que se permita la mayor precisión posible. Es decir, que dos individuos entre los que haya una determinada diferencia mínima en la duración de su vida se puedan tratar como diferentes. Por ejemplo, en vacuno lechero, si nuestra precisión es de un día, un animal que haya vivido 249 días será diferente de otro que

haya vivido 250. Pero si nuestra precisión es de lactaciones completadas, y ambos están en su primera lactación, se tratarán igual.

Una vez definida la precisión de nuestra medida de longevidad habrá que hacer frente a las peculiaridades de los datos de longevidad. La principal característica de estos datos es que pueden estar sujetos a censura, es decir, que cuando llegue el momento de hacer el análisis, de algunos animales sólo sabremos que su medida para la longevidad va a ser mayor que un determinado valor. Unas definiciones del carácter abordan este problema modelando, no directamente la longitud de la vida productiva, sino la situación del animal (vivo o muerto) en diferentes puntos de observación. Otras mantienen la definición del carácter como longitud de la vida productiva y usan diferentes técnicas para incluir la información que contienen los datos de los animales vivos. Otra peculiaridad de los datos de longevidad es que por tratarse de medidas que se llevan a cabo durante un período de tiempo más o menos largo (toda la vida de un animal) pudiera haber factores explicativos del carácter (tanto fijos como aleatorios) que a lo largo de la vida del animal cambian de nivel con el que están afectando al individuo que produce el dato (variables dependientes del tiempo). La última característica es que con ninguna de las definiciones que del carácter se han propuesto se obtienen variables respuesta normales.

De manera general hay que decir que todos los métodos y definiciones que se han propuesto tratan de afrontar el principal problema asociado con los datos de longevidad, que es la censura, pero muchos de ellos no tienen en consideración el resto de peculiaridades.

En la tesis de Chirinos (2003) se hace una revisión detallada de todos los métodos propuestos para el estudio de la longevidad. En nuestra introducción estudiaremos los métodos propuestos para el análisis de los datos de longevidad agrupándolos en función de las diferentes definiciones que del carácter se han hecho, que básicamente se pueden distribuir en dos categorías: i) “Supervivencia hasta un determinado momento” (cuando el fenotipo es la situación del animal (vivo o muerto) en diferentes momentos) y ii) “Longitud de la vida productiva” (cuando el fenotipo es la duración de la vida productiva del animal).

1.1.1.- Supervivencia hasta un determinado momento

La primera medida que se propuso de este tipo fue la *stayability*. La definición de *stayability* fue originalmente propuesta por Everett y col. (1976a, b). Este término

originalmente hacía referencia a la capacidad del animal para permanecer en el rebaño hasta una determinada edad. En trabajos posteriores se extendió la definición a la capacidad del animal para permanecer en el rebaño hasta un determinado ordinal de parto o número de lactación condicionado a que al menos el animal tuvo un primer parto (p.e. López-Serrano y col., 2000; Visscher y Goddard, 1995). Así, si el animal alcanzaba la edad, el ordinal de parto o la lactación prefijados tendría como medida fenotípica del carácter un 1 y de lo contrario tendría un 0. Con esta definición, y de una manera muy intuitiva, un semental quedará evaluado más favorablemente cuanto mayor sea la proporción de sus hijas que continúen vivas a la edad, parto o lactación prefijados.

Originalmente Everett y col. (1976a, b) propusieron para vacuno lechero observar si el animal estaba vivo o no a los 36, 48, 60, 72 y 84 meses de edad. Finalmente se fijó como umbral óptimo los 48 meses dado que a esa edad la correlación con supervivencias a edades superiores es elevada y permite una evaluación más temprana de los candidatos a la selección (Everett y col., 1976b). Un inconveniente de esta definición es la arbitrariedad a la hora de establecer la edad que los animales deben alcanzar y su pequeña precisión diferenciando animales. Por ejemplo, si dos animales mueren antes de alcanzar el umbral prefijado, ambos tendrán la misma medida fenotípica del carácter, aunque pudiera ocurrir que uno lo haga un día antes de alcanzar este umbral y el otro cinco meses antes. Por tratarse de un carácter de expresión binaria el método óptimo para la estimación de los parámetros de los modelos explicativos será algún método de análisis de datos categóricos (Gianola, 1980), aunque a nivel práctico son muchos los trabajos en los que este tipo de caracteres se estudian usando métodos estándar como los modelos lineales mixtos (Everett y col., 1976a,b; Tholen et al. 1996a,b; López-Serrano y col., 2000).

Para hacer frente a la pérdida de información, y consecuentemente a la baja precisión de la medida, inherente a la propia definición del carácter, se han propuesto varias alternativas.

Van Vleck (1980) propone mantener todos los umbrales de definición del carácter, y hacer con todos ellos un análisis categórico multicarácter. El problema asociado con este método de análisis de la longevidad es que, dada la naturaleza binaria de los caracteres, el diseño de efectos fijos puede influir en la estructura de varianzas-covarianzas, lo que se puede traducir en sesgos en las predicciones de los méritos genéticos (Van Vleck, 1980).

Madgwick y Goddard (1989) propusieron definir el carácter a través de lo que llamaron “códigos de supervivencia”. Estos códigos consisten en una serie de variables indicadoras que señalan si el animal (en su caso vacas lecheras) está vivo o no en sucesivos puntos de observación del eje del tiempo, por lo tanto, la precisión de la medida de la longevidad con esta definición vendrá dada por el intervalo de tiempo que se establezca entre las sucesivas variables indicadoras (para el caso de Madgwick y Goddard (1989) era de un año). Así, se define $S_i = 1$ si el animal sobrevive desde el año i al año $i + 1$ después del primer parto o $S_i = 0$ si no lo hace; S_i se tratará como una observación ausente si el animal fue eliminado o murió antes de i años después del primer parto. Esta aproximación comparada con la *stayability* (Everett y col., 1976a,b), además de que hace un uso más apropiado de la información, tiene otras ventajas, una de ellas, según indican los autores, es que la vida en el rebaño después de un primer parto es un carácter más relevante que la edad completa de la vaca.

En el trabajo de Madgwick y Goddard (1989) los análisis que se hacen para todos los códigos de supervivencia definidos (hasta el año 4º y hasta el año 8º, según la base de datos) usan metodología de modelo lineal mixto unicarácter. También se estudió la posibilidad de analizar los códigos con modelos no lineales y, en concreto, la correlación entre los valores genéticos para la supervivencia durante el primer año de vida (S_0) predichos usando un modelo lineal mixto y un modelo tipo umbral fue muy próxima a 1. En la discusión de este trabajo se plantea el modelo multicarácter como el más adecuado para analizar conjuntamente todos los códigos de supervivencia aunque por las elevadas necesidades de cómputo se propone el modelo de repetibilidad como una aproximación aceptable (la correlación genética entre los códigos no es realmente de uno).

Jairath y col. (1998), Visscher y Goddard (1995) y Boettcher y col. (1999) (todos estos estudios se llevan a cabo en vacuno lechero) analizan estos códigos de supervivencia mediante un análisis multicarácter. De estos tres trabajos solo en dos de ellos (Visscher y Goddard (1995) y Boettcher y col. (1999)) se estiman los componentes de varianza. Visscher y Goddard (1995) indican que para hacer la estimación de las covarianzas genéticas habrá que fijar las covarianzas ambientales a cero dado que para un animal el dato de supervivencia en una lactación (0 ó 1) sólo quedará definido si en todas las lactaciones previas de ese individuo tiene como dato un 1 (sobrevive), de esta manera se alcanzará la convergencia en el algoritmo de cómputo (REML).

El inconveniente asociado con estos análisis multicarácter es la sobreparametrización de los modelos. Para evitar este problema se ha propuesto el análisis de los datos de códigos de supervivencia a través de modelos de regresión aleatoria (Veerkamp y col., 2001), que tienen todas las ventajas de los análisis multicarácter pero el número de parámetros a estimar es mucho más reducido. En este trabajo la definición de los códigos de supervivencia se hace de manera equivalente: Si un animal en la lactación i tiene una lactación $i + 1$ ese animal habrá sobrevivido a la lactación i y su código para esa lactación será 0, si no tiene una lactación $i + 1$ habrá muerto o habrá sido eliminado y su código será 1; después de la muerte o eliminación de un animal el resto de lactaciones hasta cinco se tratan como ausentes. Schaeffer (2004) en un trabajo teórico sobre la posible aplicación de la regresión aleatoria a diferentes caracteres, entre los que se encuentra la supervivencia, propone usar una definición similar a la empleada por Madgwick y Goddard (1989) para establecer los códigos de supervivencia. En concreto Schaeffer (2004) propone observar si el animal está presente o no cada día de control lechero, así si un animal está presente le asignará como medida fenotípica un 1 y si no lo está un 0. A diferencia de Madgwick y Goddard (1989) y Veerkamp y col. (2001) una vez muerto el animal en los sucesivos puntos de observación se mantendrá el código 0 y no lo tratará como dato ausente.

Con estas definiciones del carácter el tratamiento de los datos de animales vivos en el momento del análisis es inmediato, pues en estos análisis la medida fenotípica de un individuo es precisamente si el animal está vivo o no.

Con la definición original de *stayability* (Everett y col., 1976a,b) sólo se podrían acomodar variables dependientes del tiempo si se hiciese un análisis multicarácter como el que propone Van Vleck (1980). En este tipo de análisis en cada punto de observación se pueden establecer los niveles de efectos fijos que afectan al animal que produce el dato, y para cada uno de los puntos de observación se tendrá un valor aditivo para cada animal, aunque no se han encontrado aplicaciones prácticas de este procedimiento posiblemente debido a la dificultad de manejar los efectos fijos en análisis de datos de expresión binaria y a las elevadas necesidades de cómputo que un modelo de este tipo exige. Para los códigos de supervivencia (Madgwick y Goddard, 1989), tanto si se hace un análisis multicarácter de los mismos como si se emplean técnicas de regresión aleatoria, igualmente se podrán incluir efectos fijos dependientes del tiempo, y además se pueden obtener estimas de los valores genéticos para diferentes periodos.

Con estas definiciones de longevidad es posible hacer análisis conjuntos con otros caracteres de interés, tanto si la medida fenotípica de longevidad se considera como un carácter continuo y se asume normalidad (modelo lineal mixto multicarácter), como si se considera, de una manera más realista acorde con su naturaleza discreta, como un carácter categórico. Un trabajo teórico en el que se aborda el análisis conjunto de varios tipos de caracteres -gaussianos, gaussianos con censura, binarios y categóricos con múltiples categorías de respuesta- es el de Korsgaard y col. (2003).

El principal inconveniente asociado con la definición original de Everett y col. (1976a, b) era la baja precisión de la medida del carácter. Como se ha comentado anteriormente, Schaeffer (2004) propone observar si un animal está vivo o muerto cada día de control lechero, de manera que la precisión aumentará considerablemente, pero el método de análisis que propone (regresión aleatoria) no tiene en consideración la naturaleza binaria de la definición del carácter.

1.1.2.- Longitud de la vida productiva

Otra definición del carácter es, directamente, la amplitud del intervalo desde el comienzo de la vida productiva del animal hasta su muerte o eliminación. En este caso la precisión de la medida vendrá dada por las unidades que usemos para medir ese intervalo (días, semanas, meses, lactaciones, partos, etc.).

Con esta definición, el principal inconveniente con el que nos encontramos es cómo manejar los datos censurados, datos de animales que aún están vivos en el momento de hacer el análisis o de animales que por razones distintas a la muerte o eliminación salen del sistema de control de registros productivos, pues desde un punto de vista práctico no es viable esperar a que todos los individuos mueran o sean eliminados antes de hacer el análisis. Si de lo que se trata es de hacer evaluaciones genéticas, esto provocaría grandes incrementos del intervalo generacional y además si entre los animales que producen los datos están los candidatos a la selección una vez muertos no habría individuos que seleccionar. Otra peculiaridad de esta definición del carácter es que la variable respuesta no es normal, tiene una gran cola a la derecha.

Una alternativa para el manejo de los datos censurados, cuando la definición del carácter se hace como longitud de la vida productiva, fue propuesta por VanRaden y Klaaskate (1993) en ganado vacuno lechero. Estos autores estudiaron el número de meses que una vaca está en lactación hasta los 84 meses de edad, como una medida de la vida útil de los animales. En el momento de hacer el análisis habrá animales que

estando vivos, no habrán llegado a los 84 meses de edad, o animales que son vendidos y salen del sistema de control de registros, y por lo tanto de ellos no se conocerá el número total de meses que pudieran haber estado en lactación hasta los 84 meses de edad. Para estos animales se propone predecir el número total de meses que hubieran estado en lactación hasta los 84 meses de edad mediante regresión múltiple, usando como variables independientes el rebaño-año-estación al que pertenecían originalmente, la edad al primer parto, la producción estandarizada de leche en la primera lactación, el estado de lactación, los meses que lleva en la actual lactación o periodo de secado y los meses acumulados en lactación. Una vez predicho el número de meses que una vaca viva estará en lactación, el análisis se aborda usando metodología de modelo lineal mixto, pero dando pesos diferentes a los datos completos (de animales que mueren o que llegan a los 84 meses de vida) y a los datos predichos. Los principales inconvenientes asociados con esta metodología son tres: el primero es que el coeficiente de determinación de la regresión usada para predecir los datos incompletos es muy bajo (el más alto (0.59) se obtiene cuando la predicción se hace para un animal que ya ha vivido 72 meses); el segundo es que no se pueden usar variables dependientes del tiempo para una corrección apropiada por efectos fijos; el tercero es que se asume un efecto genético no dependiente del tiempo. Como ventajas de esta aproximación están su sencillez y que se permite abordar de manera igualmente sencilla análisis conjuntos con otros caracteres de interés, mediante modelos lineales mixtos multicarácter, aunque la premisa de normalidad para los datos de longevidad (completos y predichos) quizás sea violada.

Brotherstone y col. (1997) proponen una metodología similar a la de VanRaden y Klaaskate (1993) para el manejo de los datos censurados, es decir, hacer predicciones de la longitud de la vida productiva de animales vivos en el momento de hacer el análisis. Brotherstone y col. (1997) definen la longevidad como el número de lactaciones que una vaca completa a lo largo de su vida. La predicción de los datos de individuos censurados se hace considerando las propiedades de la distribución geométrica y las probabilidades de supervivencia en sucesivas lactaciones estimadas previamente en la misma población. Con los datos completos y con los datos predichos (de animales censurados) se hace un análisis lineal multicarácter con otros caracteres de interés. En este trabajo no se pondera diferencialmente los datos predichos y los datos completos, pero dado que la varianza fenotípica de los datos predichos es menor que la de los datos completos implícitamente se da más peso (cuando se trata de predecir

valores genéticos) a los segundos que a los primeros. Al igual que con la metodología de VanRaden y Klaaskate (1993) el principal inconveniente de esta aproximación es que no se pueden modelar efectos fijos dependientes del tiempo y que el efecto genético tampoco se puede modelar como dependiente del tiempo. Además el método de análisis propuesto no tiene en consideración la no normalidad de la variable respuesta.

Otra alternativa que se ha planteado para manejar los datos censurados fue propuesta por Guo y col. (2001) y que ellos mismos aplican para el estudio genético de la longevidad en una población de cerdos. En este trabajo se define la longevidad como el número de días desde la fecha de la primera concepción hasta la fecha de eliminación o muerte. La metodología empleada en el análisis es de modelo lineal mixto, usando una aproximación bayesiana que permite, al definir la función de verosimilitud, que los datos censurados contribuyan no con su función de densidad (como hacen los datos completos) sino con la de supervivencia. El principal inconveniente de esta metodología es que no se permite el modelado de efectos (sistemáticos y genético) dependientes del tiempo y que la distribución de los datos está lejos de ser normal, que es la distribución que se asume para construir la verosimilitud. Por eso también se propone hacer una transformación logarítmica que hace que la distribución de la variable respuesta sea más próxima a la normal.

Cuando la longevidad se define como longitud de la vida productiva, otra alternativa que se plantea para hacer frente al manejo de datos censurados es usar técnicas de análisis de supervivencia. Estas técnicas de análisis serán descritas con detalle en el próximo apartado. Aquí sólo diremos que los datos censurados se pueden incluir, con la información que contienen y sin necesidad de recurrir a predicciones en las funciones de verosimilitud que se emplean para hacer inferencia. También se pueden incluir en los análisis efectos fijos dependientes del tiempo y se tiene en cuenta la naturaleza no normal del carácter. Como una ventaja más, recientemente Damgaard (2004) ha desarrollado un modelo bicarácter para estudiar conjuntamente datos de longevidad, usando técnicas de análisis de supervivencia, y de algún otro carácter de tipo gaussiano o umbral. Previamente se habían propuesto métodos aproximados para llevar a cabo la estimaciones multicarácter (Ducrocq, 2001a). Además, teóricamente, la consideración del efecto genético como dependiente del tiempo es posible (Andersen et al. 2000), aunque no se han encontrado aplicaciones prácticas que lo hagan así. El principal inconveniente de esta metodología es que los tiempos de computación son mucho más elevados que para otras técnicas de análisis. Algunas técnicas del análisis de

supervivencia pueden ser equivalentes a métodos usados para el análisis de los códigos de supervivencia (Madgwick y Goddard, 1989; Veerkamp y col., 2001).

1.2.- Análisis de supervivencia

De una manera muy general, el análisis de supervivencia es un conjunto de técnicas estadísticas que permiten el estudio de variables aleatorias positivas que miden la amplitud del intervalo entre un punto de origen y otro final (Ducrocq, 2001b).

El uso de estas técnicas no se restringe al campo de la producción animal, hay aplicaciones en otras áreas de conocimiento como epidemiología y ciencias de la salud (Sirivongs y col., 2004), economía (Kimura y Fujii, 2003), ciencias sociales (Beamonte y Bermúdez, 2003), etc. Dentro del campo de la producción animal es en el área de la mejora genética donde más extensamente se han usado, en concreto en el vacuno lechero especialmente para estudiar su longevidad, aunque también hay trabajos publicados en biotecnología (Silvestre y col., 2004b) y en etología y reproducción (Silvestre y col., 2004a).

Los puntos de origen y fin del intervalo que estemos estudiando, así como las unidades de medida de su amplitud pueden ser muy diversos dependiendo del problema que se vaya a abordar. Por ejemplo, si estamos interesados en estudiar la longevidad productiva del vacuno lechero, el punto de origen puede ser el primer parto, el punto final será el momento de la muerte o eliminación de la vaca y las unidades de medida pueden ser días, semanas, meses, años, lactaciones, etc. Si se está estudiando el tiempo que tarda en eyacular un macho cabrío durante las extracciones de semen, el punto de origen lo puede marcar el momento en el que sale al área de extracción y el punto final lo marcará el momento de la eyaculación, y la unidad de medida puede ser una unidad de tiempo como el segundo. En este último supuesto igualmente se puede estar interesado en estudiar el número de saltos previos hasta que se produce un eyaculado efectivo, en ese caso los puntos de origen y final serán los mismos que para el caso del estudio del tiempo hasta que se produce el eyaculado, mientras que la unidad de medida será el número de saltos (Silvestre y col., 2004a).

1.2.1.- Historia de la aplicación del análisis de supervivencia a la mejora genética

El primer trabajo en el que se hace una aplicación de técnicas de análisis de supervivencia a la mejora genética es el de Famula (1981). En este trabajo se propone usar la distribución exponencial para describir la vida de los animales. En este estudio se modela la esperanza de vida a través de una combinación lineal de variables explicativas todas ellas de naturaleza fija. La estimación del efecto de las covariables se lleva a cabo por métodos máximo verosímiles. A la función de verosimilitud los datos

completos contribuyen con la función de densidad, mientras que los datos censurados lo hacen con la de supervivencia. El uso de la distribución exponencial para describir los datos de longevidad en el trabajo de Famula (1981) lleva implícito una importante restricción, derivada de lo que se conoce como “falta de memoria” del modelo exponencial, es decir, que durante toda la vida de un animal su probabilidad de eliminación o muerte es la misma.

Smith y Quaas (1984) avanzaron en la aplicación del análisis de supervivencia a la mejora genética introduciendo el uso del modelo de Cox para estudiar la duración de la vida en vacuno lechero, condicional a que el animal inició una primera lactación. En este trabajo se hacen básicamente dos análisis. En el primero se plantea un modelo de efectos fijos con estratificación para el año de nacimiento de la vaca, un efecto de rebaño anidado al año de nacimiento y un efecto de macho. Con este modelo se estiman, en dos subgrupos de datos, los efectos de macho y los efectos de rebaño y a partir de estas estimas se computan las varianzas asociadas a estos efectos. Estas aproximaciones de las varianzas se emplean en un segundo análisis en el que se hace una predicción de los efectos de macho usando un modelo mixto (de fragilidad) mediante un método bayesiano empírico¹ de máximo a *posteriori*. En Smith y Allaire (1986) se hace una revisión teórica de los métodos de inferencia y de computación que en esa época había disponibles y que podían aplicarse al análisis de supervivencia. Para la estimación y predicción de efectos se discute el método de máximo a *posteriori* y cómo implementarlo a través de un algoritmo que llaman “maximización cíclica”. Para la estimación de componentes de varianza proponen varias alternativas: integración de efectos aleatorios en la función de verosimilitud; uso del algoritmo EM para llegar a estimas tipo REML de los componentes de varianza y finalmente proponen un tercer método similar al método 3° de Henderson, que es el que discuten con cierto detalle. Este método se basa en igualar las formas cuadráticas del efecto para el que se quiere estimar el componente de varianza a su esperanza y la ecuación resultante se resuelve para el componente de varianza en cuestión.

Con la tesis de Ducrocq (1987) se produce un importante avance de las técnicas de análisis de supervivencia aplicadas a la mejora genética. En este trabajo se desarrollan a nivel práctico otras propiedades del análisis de supervivencia como por ejemplo la inclusión de variables dependientes del tiempo, la aplicación de ciertos

¹ Se conocen como métodos bayesianos empíricos aquellos en los que los valores de los hiperparámetros (componentes de varianza) se calculan a partir de los propios datos en un primer análisis (Ducrocq, 1987).

métodos para valorar la bondad del ajuste de los modelos empleados y se propone un método bayesiano para la estimación de la varianza asociada a macho, asumiendo *a priori* una distribución Gamma para los efectos de macho. La elección de esta distribución como *prior* se basa en las buenas propiedades analíticas que tiene cuando se conjuga con la verosimilitud de un modelo Weibull. En concreto se puede hacer una integración analítica de los efectos de macho y estimar la varianza asociada a macho conjuntamente con el resto de efectos “fijos” y parámetros de la distribución Weibull como la moda de la distribución posterior conjunta (efectos fijos, parámetros de la distribución Weibull y varianza asociada a macho). Además, los métodos que se desarrollan en este trabajo se pueden aplicar a grandes poblaciones (miles de toros en miles de rebaños) de vacuno lechero, cosa que no era posible con la aproximación semiparamétrica que propusieron Smith y Quass (1984). Los resultados de esta tesis se publicaron en Ducrocq y col. (1988 a,b).

A partir de 1994, con la presentación del paquete estadístico “Survival Kit” (Ducrocq y Sölkner, 1994), el análisis de supervivencia aplicado a la mejora genética se convirtió en la técnica de elección para el estudio de datos de longevidad, debido a que de una manera sencilla se podían considerar con aceptable rigor todas las peculiaridades de los datos de longevidad. Este paquete estadístico implementa tanto modelos totalmente paramétricos (Weibull y Exponencial) como modelos semiparamétricos (Cox), permitiendo para ambos el modelado de variables dependientes del tiempo y el desarrollo de análisis estratificados. Cuando los modelos son de efectos fijos la estimación se lleva a cabo por máxima verosimilitud, y cuando se trata de modelos mixtos (modelos de fragilidad) la estimación es Bayesiana, permitiendo como *priors* las distribuciones Gamma, Normal y Normal Multivariante, lo que permite, en el último caso, incluir matrices de relaciones de parentesco.

Otro trabajo que hay que destacar en esta introducción histórica es el de Korsgaard y col. (1998) en el que se desarrolla un método MCMC aplicable a datos de longevidad usando modelos semiparamétricos. Con esta aproximación se permite la inclusión de tantos efectos de naturaleza aleatoria como se desee y la estimación de los componentes de varianza asociados a ellos. A pesar de que la aproximación de Korsgaard y col. (1998) permite llegar a las distribuciones marginales completas de los parámetros de interés, su uso no se ha extendido pues las necesidades de cómputo son muy elevadas. Otro punto negativo de la aproximación propuesta por Korsgaard y col. (1998) es que no se implementó la inclusión de variables explicativas dependientes del

tiempo, que es, junto con el modo de tratar los datos censurados, como se ha comentado en el punto anterior, una de las ventajas más importantes del análisis de supervivencia con respecto a otros métodos de análisis de datos de longevidad o supervivencia.

Otra referencia importante en el desarrollo y aplicación de los métodos de análisis de supervivencia a la mejora genética es la de Roxtröm y col. (2003). En este trabajo se propone sustituir la definición de la longevidad en base a toda la vida de una vaca lechera, por una definición en base a sus sucesivas lactaciones. De esta manera se pueden sustituir los modelos que originalmente se propusieron para modelar el riesgo de eliminación, en los que se incluía el número de lactación como un efecto dependiente del tiempo (ver [apartado 1.2.4.1.2.](#)), por otros en los que no es necesario incluirlo como tal sino que se incluye en la función de riesgo base (ver [apartado 1.2.4.](#)) pudiéndose incluso estratificar (ver [apartado 1.2.4.1.1.](#)) para cada lactación. Esta sustitución se hace con el objetivo de describir mejor el efecto del número de lactación. Para el nuevo modelo que estos autores proponen cada lactación es tratada como una medida de supervivencia, el punto de origen es la fecha del parto y el punto final es la fecha de eliminación de la vaca (dato no censurado) o del siguiente parto (dato censurado). Para esta definición de la longevidad los autores demuestran que la función de verosimilitud es estrictamente equivalente a la obtenida cuando se define la longevidad en base a toda la vida del animal, sin que sea necesaria la inclusión de un efecto permanente de animal a lo largo de todas las lactaciones de una vaca. El resultado más interesante de este trabajo es que se obtiene un mejor modelado de los datos cuando se usa esta definición del carácter en lugar de la tradicional. Recientemente, Ducrocq (2005) ha publicado un trabajo en el que se extiende el concepto de longevidad definida en base a lactación, a longevidad definida en base a estado de lactación dentro de lactación y de año. De esta manera la función de riesgo base se define para cada uno de los estados de lactación que el autor define (parto-270 días; 271-380 días; 381-secado; secado-parto) en cada una de las lactaciones del animal (1^a - $\geq 5^a$) y en cada uno de los diferentes años de estudio (1988-2003). De esta manera el número de parámetros a estimar aumenta considerablemente pero el número de registros elementales (registros en la base de datos después de considerar los cambios en las covariables dependientes del tiempo) disminuye. Este autor aplica este nuevo modelo a la evaluación genética de sementales en Francia y lo somete a dos tests de validación. En el primero calcula la tendencia genética y observa que es casi plana, mientras que con el modelo de longevidad en base a la vida completa del animal es claramente creciente. El resultado con el modelo de longevidad en base a

estado de lactación dentro de lactación y de año, tiene más sentido dada la fuerte selección que se ha hecho para producción y la desfavorable correlación genética entre longevidad funcional y producción. Además la estima del coeficiente de regresión de la tendencia genética está dentro del intervalo definido por Interbull para considerar al modelo como válido. El segundo test de validación consiste en determinar si la función de supervivencia (estimada usando el estimador de Kaplan-Meier (ver [apartado 1.2.3.2.1](#))) de hijas de toros con diferente valor genético (estas hijas no se usaron en la evaluación de los toros) se corresponde con el mayor o menor valor genético de los toros. Se comprueba que, efectivamente, la función de supervivencia de las hijas de los toros con el mayor valor genético es mejor que la de las hijas de los toros con menor valor genético.

1.2.2.- Definiciones generales

Al principio de este apartado se dio una definición general de las variables de que se ocupa el análisis de supervivencia. Aquí la repetiremos: el análisis de supervivencia se ocupa del estudio de variables aleatorias positivas que miden la amplitud del intervalo entre un punto inicial y otro final. En el contexto del estudio de la longevidad productiva de los animales el punto de inicio lo marca el inicio de la vida productiva (primer parto, primera cubrición fértil, primer intento de cubrición, etc.) y el final de la vida lo determinará la muerte o eliminación del animal. Las unidades de medida serán unidades de tiempo (días, semanas, meses, años, etc.) o alguna otra unidad aunque no sea de tiempo real (lactaciones, partos, etc.).

1.2.2.1.- Censura

Este término se ha usado anteriormente en esta introducción, aunque no se ha definido de manera explícita. A nuestros efectos, se dice que hay censura cuando entre los datos hay unidades experimentales que no llegan a sufrir el acontecimiento que marca el fin del intervalo que estamos estudiando (Ducrocq, 2001b). En el contexto del estudio de la longevidad se tratará el dato de un individuo como censurado si en el momento de hacer el análisis continúa vivo o ha experimentado otro acontecimiento que es incompatible con el conocimiento de su estado, vivo o muerto (p.e.: sale del sistema de control de registros).

Existen otros tipos de censura, pero la descrita en el párrafo anterior es, con mucho, la más frecuente en los análisis de longevidad, y se le llama específicamente censura por la derecha.

Los datos censurados son datos con información parcial, no se sabe la duración de la vida de los animales censurados pero si se sabe que al menos es mayor que el tiempo desde el inicio de la vida productiva hasta el momento de observación.

La consideración de estos datos como datos completos, datos de los que se sabe cuando se inicia y cuando termina la vida productiva, implicará subestimaciones de la supervivencia media de la población y su eliminación puede producir igualmente sesgos en las estimaciones (según el valor de los datos censurados pueden producir sobreestimaciones o subestimaciones), aunque siempre supondrá una pérdida de información.

Por lo tanto, el dato de longevidad del animal i va a estar formado por un par de variables aleatorias (y_i, δ_i) , donde $y_i = \min(t_i, c_i)$, siendo t_i y c_i a su vez las realizaciones de dos variables aleatorias que representan respectivamente los posibles tiempos de muerte o de censura y δ_i es una variable indicadora que toma el valor 1 si $t_i \leq c_i$ (dato completo) ó 0 si $t_i > c_i$ (dato censurado). Una asunción general que hay implícita en todas las técnicas de análisis de supervivencia es que la censura es aleatoria, esto quiere decir que la variable T (tiempos de muerte) será independiente de la variable C (tiempos de censura). En muchos casos esta condición no se cumple, sin embargo es suficiente con que el mecanismo de censura sea independiente para que las funciones de verosimilitud, a través de las que se lleva a cabo la inferencia, que serán descritas posteriormente, sean válidas. La condición necesaria para que un mecanismo de censura sea independiente es que los individuos que se censuran en el momento t formen una muestra representativa de los individuos a riesgo en ese momento, es decir, que tienen la misma probabilidad de sufrir el acontecimiento de interés (en el caso del estudio de la longevidad la muerte o eliminación) que aquellos que no se censuran (Kalbfleisch y Prentice, 1980).

1.2.2.2.- Descripción de los datos de longevidad

Si asumimos que no hay censura en una muestra de datos de longevidad, estos datos los representará la variable T (variable no-negativa) y ésta se podrá describir a través de una serie de funciones, que están relacionadas como a continuación se verá:

1.- Función de supervivencia: $S(t) = \Pr[T \geq t]$

Describe la probabilidad de que la muerte se produzca después del tiempo t . Representa la fracción de individuos que continúan vivos en el tiempo t .

2.- Función de distribución: $F(t) = 1 - S(t) = \Pr[T < t]$

Es el complemento de la función de supervivencia, y por lo tanto representa la fracción de individuos muertos hasta el tiempo t . No es específica del análisis de supervivencia.

3.- Función de densidad: $f(t) = \frac{\lim_{dt \rightarrow 0} \Pr[t \leq T \leq t + dt]}{dt}$

Al igual que la anterior no es específica del análisis de supervivencia y en el contexto del análisis de supervivencia describe el riesgo instantáneo de muerte en el tiempo t . Sus relaciones con la función de distribución y de supervivencia son:

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt}$$

4.- Función de riesgo: $h(t) = \frac{\lim_{dt \rightarrow 0} \Pr[t \leq T \leq t + dt | T \geq t]}{dt}$

Describe el riesgo instantáneo de muerte en el tiempo t , condicional a la supervivencia hasta ese tiempo t . Es el límite del cociente entre una probabilidad condicional y un intervalo de tiempo. Este cociente será siempre positivo y puede ser mayor que 1. Esta función mide el riesgo de morir en cada tiempo t . Se relaciona con la función de supervivencia y de densidad a través de esta ecuación:

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d \log(S(t))}{dt}$$

5.- Función de riesgo acumulado: $H(t) = \int_0^t h(u) du$

Es la función de riesgo integrada hasta el tiempo t . La función de supervivencia se relaciona con ella de la siguiente manera:

$$S(t) = \exp\{-H(t)\}$$

Como se ve todas estas funciones están relacionadas entre si y el conocimiento de una de ellas implica la definición inmediata de todas las demás.

1.2.3.- Estimación de las funciones que describen los datos de longevidad

Básicamente se presentan dos alternativas para estimar las funciones que describen los datos de supervivencia de una población. Estas dos alternativas son: las que utilizan métodos paramétricos y las que utilizan métodos no paramétricos.

1.2.3.1.- Métodos paramétricos

Consisten en asumir algún modelo probabilístico que creamos se ajusta más o menos adecuadamente a los datos de que se dispone y estimar los parámetros que definen ese modelo probabilístico. Cualquiera de los modelos probabilísticos que se propongan ha de tener en consideración que la variable objeto de estudio es no-negativa.

Los modelos más extendidos en la descripción paramétrica de las funciones que definen los datos de longevidad son: Exponencial, Weibull, Log-Normal, Gamma, Gamma Generalizado, Log-Logistic y F Generalizado (Kalbfleisch y Prentice, 1980). Muchos de estos modelos son casos particulares de otros, por ejemplo el modelo Exponencial es un caso particular del modelo Weibull.

1.2.3.1.1.- Estimación bajo modelos paramétricos

Una vez escogido algún modelo probabilístico adecuado para describir los datos de longevidad disponibles, la estimación de los parámetros del modelo exigirá el cálculo de su función de verosimilitud. En los datos posiblemente habrá completos y censurados (por la derecha). A continuación se detallará cómo contribuyen cada uno de estos tipos de datos a la función de verosimilitud, $L(\beta)$.

Recordemos que los datos observados van a ser (y_i, δ_i) , donde $y_i = \min(t_i, c_i)$, siendo t_i y c_i las realizaciones de dos variables aleatorias que representan respectivamente los posibles tiempos de muerte o de censura y δ_i es una variable indicadora que toma el valor 1 si $t_i \leq c_i$ (dato completo) ó 0 si $t_i > c_i$ (dato censurado). Las variables “tiempos de censura” y “tiempos de muerte” se podrán describir a través de las funciones de densidad, $g(c)$ y $f(t)$, y de supervivencia, $G(c) = \Pr[C \geq c]$ y $S(t) = \Pr[T \geq t]$.

Datos completos:

$$\begin{aligned} L_i &= \Pr[y_i \in (y_i, y_i + dt), \delta_i = 1] / dt \\ &= \Pr[t_i \in (y_i, y_i + dt), c_i \geq y_i] / dt \end{aligned}$$

Como consecuencia de la censura aleatoria

$$\begin{aligned} L_i &= \Pr[t_i \in (y_i, y_i + dt)] / dt \times \Pr[c_i \geq y_i] \\ &= f(y_i) \times G(y_i) \end{aligned}$$

Datos censurados por la derecha:

$$\begin{aligned} L_i &= \Pr[y_i \in (y_i, y_i + dt), \delta_i = 0] / dt \\ &= \Pr[c_i \in (y_i, y_i + dt), t_i \geq y_i] / dt \end{aligned}$$

Como consecuencia de la censura aleatoria

$$\begin{aligned} L_i &= \Pr[c_i \in (y_i, y_i + dt)] / dt \times \Pr[t_i \geq y_i] \\ &= g(y_i) \times S(y_i) \end{aligned}$$

Por lo tanto si llamamos β al conjunto de parámetros que describen al modelo probabilístico escogido, la función de verosimilitud se puede expresar:

$$L(\beta) = \prod_{i \in \{com.\}} [f(y_i)G(y_i)] \prod_{i \in \{cen.\}} [g(y_i)S(y_i)]$$

Donde $\{com.\}$ y $\{cen.\}$ representan los conjuntos de datos completos y censurados.

Asumiendo que la censura es aleatoria, la función de densidad y de supervivencia para los tiempos de censura no dependerán de β , por lo que la función de verosimilitud será proporcional a:

$$\begin{aligned} L(\beta) &\propto \prod_{i \in \{com.\}} [f(y_i)] \prod_{i \in \{cen.\}} [S(y_i)] \\ L(\beta) &\propto \prod_i [f(y_i)]^{\delta_i} [S(y_i)]^{1-\delta_i} \end{aligned}$$

Y dada la relación entre $f(y_i)$ y $h(y_i)$ esta función de verosimilitud se puede expresar como:

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

Para hacer estimaciones máximo verosímiles habrá que maximizar esta función. El máximo de esta función se encontrará en el mismo punto que el de su logaritmo por lo que se tomará el logaritmo de la función para hacer el álgebra más sencilla.

$$\log L(\beta) = cte. + \sum_{i \in \{com.\}} \log h(y_i) + \sum_i \log S(y_i)$$

Las estimas máximo verosímiles de β ($\hat{\beta}$) serán asintóticamente normales con media β y matriz de varianzas-covarianzas V , siendo V ,

$$V = \left[-\frac{d^2 \log L(\beta)}{d\beta^2} \right]_{\beta=\hat{\beta}}^{-1}$$

Para llevar a cabo la maximización se llegará a un sistema de ecuaciones no lineales que se puede resolver usando algún método iterativo, como por ejemplo el método de Newton-Raphson descrito en Klein y Moeschberger (1997).

1.2.3.2.- Métodos no paramétricos

Dentro de la metodología del análisis de supervivencia las técnicas no paramétricas son muy populares, pues permiten obtener estimas de las funciones que describen los datos de longevidad sin necesidad de asumir ningún modelo probabilístico concreto. Generalmente se usan como métodos descriptivos de los datos de que se dispone y permiten decidir si algún modelo paramétrico es adecuado para describir los datos recogidos. También existen tests de hipótesis no paramétricos para estimar la diferencia en supervivencia entre varios grupos de individuos dentro de la misma base de datos. Todos estos métodos no paramétricos se encuentran desarrollados con detalle en textos como Kalbfleisch y Prentice, (1980), Collett (2003), Klein y Moeschberger (1997) y Ducrocq (2001b).

1.2.3.2.1.- Estimador de Kaplan-Meier de la función de supervivencia y otros estimadores relacionados

Se basa en la descripción de la definición probabilística de la función de supervivencia como un producto de probabilidades condicionales:

$$\begin{aligned} S(t) &= \Pr[T \geq t] \\ &= \Pr[T > T_{[1]} | T > 0] \times \Pr[T > T_{[2]} | T > T_{[1]}] \times \dots \\ &\dots \times \Pr[T > T_{[j]} | T > T_{[j-1]}] \times \dots \times \Pr[T > T_{[K]} | T > T_{[K-1]}] \end{aligned}$$

$T_{[i]}$ para $i = 1, 2, \dots, K$ son los tiempos ordenados en los que se producen muertes (tiempos de fallo), siendo $T_{[K]}$ el último tiempo de muerte. Un estimador de cada una de estas probabilidades condicionales será, para el caso general i :

$$\frac{\text{n}^\circ \text{ de individuos vivos justo después de } T_{[i]}}{\text{n}^\circ \text{ de individuos vivos justo antes de } T_{[i]}} = \frac{n_i - d_i}{n_i}$$

Donde n_i son los individuos presentes justo antes de $T_{[i]}$ y d_i son los individuos que mueren en $T_{[i]}$. Si hay individuos censurados estos se tienen en consideración pues no están presentes en el grupo a riesgo (que pueden morir) n_i . Definidas de esta manera las probabilidades condicionales, el estimador de Kaplan-Meier de la función de supervivencia será:

$$\hat{S}_{KM}(t) = \prod_{i: T_{[i]} < t} \left(\frac{n_i - d_i}{n_i} \right)$$

La obtención de la fórmula para la varianza del estimador de Kaplan-Meier se basa en la asunción de que el número de individuos que sobreviven en un tiempo $T_{[i]}$ ($n_i - d_i$) sigue una distribución binomial con parámetros n_i y p_i , siendo p_i la verdadera probabilidad de supervivencia. De esta manera la varianza de $n_i - d_i$ será $n_i p_i (1 - p_i)$.

Como la estima de la probabilidad de supervivencia en un tiempo $T_{[i]}$ es $\hat{p}_i = \frac{n_i - d_i}{n_i}$, su

varianza se puede aproximar por $Var[\hat{p}_i] \approx \frac{\hat{p}_i(1 - \hat{p}_i)}{n_i}$. La varianza del logaritmo del

estimador de Kaplan-Meier de la función de supervivencia será $Var[\log(\hat{S}_{KM}(t))] = \sum_{i: T_{[i]} < t} Var[\log(\hat{p}_i)]$. Conocida $Var[\hat{p}_i]$, $Var[\log(\hat{p}_i)]$ se puede

aproximar usando el “método delta”:

$$Var[\log(\hat{p}_i)] \approx \frac{d_i}{n_i(n_i - d_i)}$$

La varianza del logaritmo del estimador de Kaplan-Meier será $Var\left[\log\left(\hat{S}_{KM}(t)\right)\right] \approx \sum_{i|T_{[i]} < t} \frac{d_i}{n_i(n_i - d_i)}$. Para llegar a la varianza del estimador de Kaplan-

Meier habrá que aplicar una segunda vez el “método delta”:

$$Var\left[\log\left(\hat{S}_{KM}(t)\right)\right] \approx \frac{1}{\left[\hat{S}_{KM}(t)\right]^2} Var\left[\hat{S}_{KM}(t)\right]$$

por lo que la varianza del estimador de Kaplan-Meier será:

$$Var\left[\hat{S}_{KM}(t)\right] \approx \left[\hat{S}_{KM}(t)\right]^2 \sum_{i|T_{[i]} < t} \frac{d_i}{n_i(n_i - d_i)}$$

A esta última ecuación se la conoce como fórmula de Greenwood y permite construir intervalos de confianza para las funciones de supervivencia obtenidas mediante el estimador de Kaplan-Meier.

Se han propuesto otros métodos para derivar el estimador de Kaplan-Meier, en concreto se puede llegar a la misma ecuación usando métodos máximo verosímiles.

Por la relación entre la función de supervivencia y la de riesgo ($h(t) = -\frac{d \log(S(t))}{dt}$), se puede llegar a un estimador no paramétrico de la función de

riesgo. Para el tiempo i será: $\hat{h}_i = \frac{d_i}{n_i}$. A partir de este estimador y dada su relación con

la función de riesgo acumulado ($H(t) = \int_0^t h(u) du$) un estimador de ésta será:

$$\hat{H}_N = \sum_{i|T_{[i]} < t} \frac{d_i}{n_i}$$

A este estimador se le conoce como estimador de Nelson. Si en lugar de la relación entre la función de riesgo y la de riesgo acumulado usamos la relación entre la función de supervivencia y la de riesgo acumulado ($H(t) = -\log S(t)$) y sustituimos la función de supervivencia por su estimador de Kaplan-Meier llegaremos a lo que se conoce como estimador de Peterson de la función de riesgo acumulado:

$$\hat{H}_P = \sum_{i|T_{[i]} < t} -\log\left(1 - \frac{d_i}{n_i}\right)$$

Ambos estimadores son asintóticamente equivalentes pues para n_i lo suficientemente grandes $-\log(1 - \frac{d_i}{n_i}) \approx \frac{d_i}{n_i}$.

1.2.3.2.2.-Test de hipótesis no paramétricos

Hasta ahora hemos visto cómo en una población homogénea se pueden estimar las funciones que describen los datos de longevidad. Ahora el problema que se plantea es diferente, se van a describir una serie de tests de hipótesis que permiten inferir si la supervivencia o cualquier otra función usada para describir los datos de longevidad, es diferente entre subpoblaciones afectadas por diferentes niveles de un mismo factor. Como se verá más adelante para responder a esta pregunta se pueden usar métodos paramétricos, semiparamétricos o, como vamos a ver ahora, métodos no paramétricos.

Para el caso general en que haya K niveles o subpoblaciones dentro de la población, el test que se planteará será:

$$H_0 : h_1(t) = h_2(t) = \dots = h_K(t) \text{ para todo } t \leq \tau$$
$$H_1 : \text{Al menos uno de los } h_j(t) \text{ es diferente para algún } t \leq \tau$$

Donde τ es el tiempo de muerte más grande en el estudio. La hipótesis nula será rechazada si, al menos, una de las poblaciones difiere de las otras en algún tiempo.

Los tiempos de muerte se ordenarán de manera creciente para todas las subpoblaciones en conjunto ($t_1 < t_2 < t_3 \dots < t_D = \tau$). En el tiempo t_i se observarán d_{ij} muertes en la subpoblación j , de n_{ij} individuos a riesgo en esa subpoblación.

Para llevar a cabo el test se harán comparaciones ponderadas del riesgo estimado en la j -ésima subpoblación bajo la hipótesis nula y bajo la hipótesis alternativa, basándonos en el estimador de Nelson. Si la hipótesis nula es cierta entonces un estimador del riesgo esperado en la j -ésima subpoblación, en un tiempo i , será $\frac{d_i}{n_i}$ (la j -ésima subpoblación se comporta igual que el conjunto de la población). El estadístico será:

$$Z_j(\tau) = \sum_{i=1}^D W_j(t_i) \left\{ \frac{d_{ij}}{n_{ij}} - \frac{d_i}{n_i} \right\}, j = 1, \dots, K$$

En este estadístico $W_j(t_i)$ es una función que pondera los diferentes tiempos, es decir que permite dar más o menos peso a los diferentes puntos de muerte y D es el número total de tiempos de muerte observados.

Si todos los $Z_j(\tau)$ están cercanos a cero habrá poca evidencia para rechazar la hipótesis nula, mientras que si uno de los $Z_j(\tau)$ está lejos de cero esta subpoblación tiene un riesgo diferente del esperado bajo la hipótesis nula.

En la forma del test estadístico visto hasta ahora se permite una función de ponderación diferente para cada subpoblación. En la práctica esta función de ponderación suele ser $W_j(t_i) = n_{ij}W(t_i)$, de tal manera que la función $W(t_i)$ es común para todos los grupos que se están comparando. Por lo tanto el estadístico tendrá la forma:

$$Z_j(\tau) = \sum_{i=1}^D W(t_i) \left\{ d_{ij} - n_{ij} \frac{d_i}{n_i} \right\}, j = 1, \dots, K$$

Las varianzas y covarianzas para este estadístico serán:

$$\begin{aligned} \text{Var}[Z_j(\tau)] &= \sum_{i=1}^D W(t_i)^2 \frac{n_{ij}}{n_i} \left(1 - \frac{n_{ij}}{n_i} \right) \left(\frac{n_i - d_i}{n_i - 1} \right) d_i, j = 1, \dots, K \\ \text{Cov}[Z_j(\tau), Z_g(\tau)] &= \sum_{i=1}^D W(t_i)^2 \frac{n_{ij}}{n_i} \frac{n_{ig}}{n_i} \left(\frac{n_i - d_i}{n_i - 1} \right) d_i, g \neq j \end{aligned}$$

En la ecuación de las varianzas el término $\frac{n_{ij}}{n_i} \left(1 - \frac{n_{ij}}{n_i} \right) d_i$, y en la de las covarianzas el

término $\frac{n_{ij}}{n_i} \frac{n_{ig}}{n_i} d_i$ proceden de la varianza y covarianza de una distribución multinomial

con parámetros $N = d_i$ y $p_j = \frac{n_{ij}}{n_i}, j = 1, \dots, K$, mientras que los términos $\frac{n_i - d_i}{n_i - 1}$ son

correcciones para cuando en un tiempo de fallo varios individuos mueren, si d_i es igual a 1 el término de corrección se simplificará también a 1.

Los componentes del vector $(Z_1(\tau), \dots, Z_K(\tau))$ son linealmente dependientes debido a que $\sum_{j=1}^K Z_j(\tau) = 0$, por lo tanto el test estadístico se construirá eligiendo

$K - 1$ $Z_j(\tau)$ y la correspondiente matriz de varianzas-covarianzas (Σ) de dimensión

$(K - 1 \times K - 1)$. La forma cuadrática $(Z_1(\tau), \dots, Z_{K-1}(\tau))\Sigma^{-1}(Z_1(\tau), \dots, Z_{K-1}(\tau))'$ se distribuirá, bajo la hipótesis nula, de acuerdo a una Chi-cuadrado con $K - 1$ grados de libertad. Por lo que ya tenemos una forma de cuantificar lo diferentes de cero que son los $Z_j(\tau)$.

El log-rank test es, con mucho, el test más popular de todos los existentes, para este test la función de ponderación ($W(t_i)$) es 1 para todos los tiempos, de manera que no se da más peso a unas regiones del tiempo que a otras. Este test tiene una potencia óptima cuando los ratios de riesgo entre las K subpoblaciones son proporcionales a lo largo del tiempo. Cuando la función de ponderación es n_i , al test estadístico se le llama test de Wilcoxon. Este test da más peso a regiones de tiempo donde hay más individuos a riesgo, que será al principio del periodo de estudio.

1.2.4.-Modelos de regresión

Conceptualmente son una generalización de los métodos no paramétricos descritos en el punto anterior, en el sentido de que permiten estudiar el efecto de diferentes factores sobre las funciones que describen los datos de longevidad.

La manera más común de presentar estos modelos es a través de la función de riesgo y básicamente hay tres alternativas: i) modelos de riesgos proporcionales; ii) modelos de tiempo de fallo acelerado y iii) modelos aditivos. Los que se usan más comúnmente son los primeros.

La ecuación general de los modelos de riesgos proporcionales es:

$$h(t | x') = h_0(t)c(x'\beta)$$

Donde $h_0(t)$, se conoce como función de riesgo base, describe el proceso de envejecimiento de la población y se puede definir de manera paramétrica o no, $c(x'\beta)$ es una función no negativa del efecto (β) de las covariables (x'). La elección más común para esta función es $\exp\{x'\beta\}$. Por lo que, en estos modelos, las funciones de riesgo y de supervivencia necesarias para construir la función de verosimilitud, con la que se abordará la estimación, serán:

$$h(t | x') = h_0(t) \exp\{x'\beta\}$$
$$S(t | x') = S_0(t)^{\exp\{x'\beta\}}$$

A estos modelos se los conoce como de riesgos proporcionales pues para dos individuos i y j con covariables x'_1 y x'_2 , el ratio de riesgo se mantendrá constante a lo largo de todo el periodo de tiempo.

$$\frac{h_i(t | x'_1) = h_0(t) \exp\{x'_1\beta\}}{h_j(t | x'_2) = h_0(t) \exp\{x'_2\beta\}} = \frac{\exp\{x'_1\beta\}}{\exp\{x'_2\beta\}} = \exp\{(x'_1 - x'_2)\beta\}$$

Una manera alternativa de presentar los modelos de riesgos proporcionales es hacerlo a través de su modelo log-lineal equivalente. Para diferentes modelos paramétricos la forma general será:

$$y = \log(t) = \mu + x'\beta + \sigma\omega$$

Donde x' es un vector de variables independientes (continuas o categóricas), β es el vector de los coeficientes de regresión, ω es un término residual y σ será un parámetro de escala particular para el modelo probabilístico escogido.

De manera más general, para cualquier forma de la función de riesgo base del modelo de riesgos proporcionales, igualmente se podrá obtener un modelo log-lineal equivalente.

El riesgo para el animal i condicionado a las covariables que le afectan vendrá dado por:

$$h_i(t | x'_i) = h_0(t) \exp\{x'_i\beta\}$$

La función de riesgo acumulado para ese individuo será:

$$H_i(t | x'_i) = H_0(t) \exp\{x'_i\beta\}$$

Y la de supervivencia:

$$S_i(t | x'_i) = \exp\{-H_i(t | x'_i)\}$$

$S_i(t_i | x'_i) = \exp\{-H_i(t_i | x'_i)\}$ (siendo t_i la realización para el individuo i de la variable tiempo de fallo) se distribuye de manera uniforme en el intervalo $[0,1]$ (Andersen et al., 2000), por lo que el riesgo acumulado $y_i = H_i(t_i | x'_i)$ lo hará de manera exponencial con parámetro 1, y por lo tanto su logaritmo lo hará siguiendo una distribución de valor extremo. Si a ese logaritmo lo llamamos ε_i :

$$\varepsilon_i = \log(y_i) = \log(H_i(t_i | x'_i)) = \log(H_0(t_i)) + x'_i\beta$$

Reordenando términos tendremos:

$$\log(H_0(t_i)) = -x'_i\beta + \varepsilon_i$$

Donde, como hemos dicho, ε_i es una variable que se distribuye de acuerdo a un modelo de valor extremo y el modelo de riesgo proporcionales es equivalente a un modelo log-lineal en la escala del $\log(H_0(\cdot))$.

En los modelos de tiempo de fallo acelerado no sólo se asume que las covariables actúan de manera multiplicativa sobre la función de riesgo base sino que también provocan un cambio en la escala del tiempo. Su ecuación es:

$$h(t | x') = h_0(t \exp\{x'\beta\}) \exp\{x'\beta\}$$
$$h(t | x') = h_0(t^*) \exp\{x'\beta\}$$

Los modelos de regresión de Weibull son un caso muy particular pues a la vez son modelos de riesgos proporcionales y de tiempo de fallo acelerado (Ducrocq, 2001b).

Otro grupo de modelos de regresión son los modelos de riesgo aditivo, en los que su ecuación general es:

$$h(t | x') = h_0(t) + x'\beta$$

Donde $h_0(t)$ es, al igual que antes, una función de riesgo base común para toda la población que describe el proceso de envejecimiento de la misma, y $x'\beta$ es el efecto de las covariables que es específico para cada individuo o grupos de individuos. Los valores de los parámetros de regresión pueden ser tanto positivos como negativos, pero hay que imponer la restricción de que $h(t | x')$ ha de ser positiva.

Los modelos de riesgo proporcionales son los más comúnmente empleados en el estudio de la longevidad, pues el riesgo de que se produzca la muerte de un individuo va a estar determinado más por la acción súbita de un efecto, que por la acción aditiva y acumulativa de factores, como se asume con el modelo de riesgos aditivos (Chirinos, 2003). De los modelos de tiempo de fallo acelerado no se han encontrado aplicaciones directas para estudiar la longevidad de los animales de producción, si se exceptúa el

modelo de Weibull. Ya se ha comentado que el modelo de Weibull se puede interpretar como un modelo de tiempo de fallo acelerado, y es este modelo el más comúnmente utilizado en el estudio de la longevidad en las especies ganaderas

1.2.4.1.- Extensión de los modelos de riesgos proporcionales

Una limitación de los modelos de riesgos proporcionales es que, a veces, o no es posible o es incorrecto asumir que entre dos individuos cogidos al azar de una población el ratio de riesgo se mantendrá constante a lo largo de todo el periodo de tiempo. Esto puede ser debido a que los individuos pertenezcan a dos diferentes subpoblaciones (estratos) que tienen diferentes funciones de riesgo base o a que existen variables explicativas que pueden cambiar de nivel con el tiempo.

1.2.4.1.1.- Estratificación

Asumamos dos individuos i y j , que pertenecen a dos estratos diferentes 1 y 2, y por tanto la función de riesgo base de cada uno de ellos será diferente. El ratio de riesgo entre ellos será:

$$\frac{h_i(t | x'_i) = h_{0,1}(t) \exp\{x'_i \beta\}}{h_j(t | x'_j) = h_{0,2}(t) \exp\{x'_j \beta\}} \neq \text{cte. (depende del tiempo)}$$

En este caso, la proporcionalidad entre los riesgos sólo se mantendría si los dos individuos perteneciesen al mismo estrato.

1.2.4.1.2.- Covariables dependientes del tiempo

Igualmente asumamos dos individuos i y j , afectados por diferentes niveles de las covariables en función del tiempo, su ratio de riesgo será:

$$\frac{h_i(t | x'_i(t)) = h_0(t) \exp\{x'_i(t) \beta\}}{h_j(t | x'_j(t)) = h_0(t) \exp\{x'_j(t) \beta\}} = \exp\{(x'_i(t) - x'_j(t)) \beta\} \neq \text{cte. (depende del tiempo)}$$

La función que se asume para describir el cambio de las covariables con el tiempo es generalmente una función de escalera en la que el valor de las covariables se mantiene constante entre dos puntos de cambio de nivel de las mismas (*piecewise constant function*). Por lo tanto este ratio de riesgo solo será constante durante el periodo de tiempo en el que no se produzcan cambios de nivel en las covariables que afectan a cualquiera de los dos individuos.

Por lo tanto, con la consideración de la estratificación y de las variables dependientes del tiempo la ecuación más general del modelo de riesgos proporcionales para el individuo i será:

$$h_i(t | x'_i(t)) = h_{0,s}(t) \exp\{x'_i(t)\beta\}$$

Donde s representa el estrato al que pertenece el individuo i .

En función de cómo se defina la función de riesgo base para los diferentes estratos tendremos las dos grandes familias de los modelos de riesgos proporcionales, paramétricos o semiparamétricos.

1.2.4.2.- Modelos de riesgos proporcionales paramétricos

En estos modelos la función de riesgo base se define a través de algún modelo probabilístico. Con mucho, el modelo Weibull es el más empleado en el contexto de la mejora genética para estudiar la longevidad de los animales, por lo tanto será para éste para el que explicitaremos las funciones que lo describen y que son necesarias para construir la función de verosimilitud, para el caso en que no haya estratificación ni covariables dependientes del tiempo.

$$h(y | x') = \lambda \rho (\lambda y)^{\rho-1} \exp\{x'\beta\} = \rho y^{\rho-1} \exp\{\rho \log \lambda + x'\beta\}$$
$$S(y | x') = \exp\left\{-\int_0^y \rho u^{\rho-1} \exp\{\rho \log \lambda + x'\beta\} du\right\} = \exp\{-y^\rho \exp\{\rho \log \lambda + x'\beta\}\}$$

En estas ecuaciones la variable y tiene el mismo sentido que la definición que de ella se hizo en los puntos [1.2.2.1](#) y [1.2.3.1.1](#).

La estimación bajo este modelo se lleva a cabo por máxima verosimilitud (igual que se especificó en el punto [1.2.3.1.1](#)). Entre los parámetros a estimar estarán los que describen la función de riesgo base y los efectos de las covariables.

1.2.4.3.- Modelos de riesgos proporcionales semiparamétricos

En este tipo de modelos la función de riesgo base no se describe por ningún modelo probabilístico estándar, sino que, o bien se mantiene sin definir (modelo de Cox) o se define a través de una función que no es propia de ningún modelo probabilístico estándar.

1.2.4.3.1.- Modelo de Cox

Este modelo fue originalmente propuesto por Cox (1972) y, como ya se ha comentado, en él la función de riesgo base se mantiene indefinida. Esto es posible pues la estimación de los parámetros se hace a través de lo que se llama función de verosimilitud parcial, que solo depende del efecto de las covariables del modelo, y no de la función de riesgo base.

El desarrollo de la función de verosimilitud parcial (Ducrocq, 2001b) se hace por un procedimiento similar al detallado anteriormente para el estimador de Kaplan-Meier de la función de supervivencia.

Supongamos que hay K tiempos de muerte y que están ordenados de menor a mayor, de tal manera que $T_{[1]} < T_{[2]} < \dots < T_{[i]} \dots < T_{[K]}$ y que en cada uno de esos tiempos de muerte sólo muere un individuo. De esta manera, la probabilidad de que el individuo j muera el en tiempo $T_{[i]}$, condicional a que hasta ese momento ha estado vivo (pertenece al grupo a riesgo, $R(T_{[i]})$) y a que en ese tiempo se ha de producir una muerte será:

$$\Pr \left[j \text{ muera en el tiempo } T_{[i]} \mid j \in R(T_{[i]}), \text{ tiene que ocurrir una muerte en } T_{[i]} \right] = \frac{\Pr \left[j \text{ muera en el tiempo } T_{[i]} \mid j \in R(T_{[i]}) \right]}{\sum_{h \in R(T_{[i]})} \Pr \left[h \text{ muera en el tiempo } T_{[i]} \mid h \in R(T_{[i]}) \right]}$$

La extensión de esta ecuación para el caso continuo será:

$$\lim_{dt \rightarrow 0} \frac{\Pr \left[j \text{ muera en } (T_{[i]}, T_{[i]} + dt) \mid j \in R(T_{[i]}) \right]}{\sum_{h \in R(T_{[i]})} \Pr \left[h \text{ muera en } (T_{[i]}, T_{[i]} + dt) \mid h \in R(T_{[i]}) \right]}$$

Por la definición de función de riesgo las probabilidades de este límite serán:

$$\lim_{dt \rightarrow 0} \frac{h_0(T_{[i]}) \exp \{x'_j \beta\} dt}{\sum_{h \in R(T_{[i]})} h_0(T_{[i]}) \exp \{x'_h \beta\} dt}$$

Por lo que la probabilidad de que el individuo j muera en el tiempo i , condicional a que hasta ese momento ha estado vivo y a que en ese tiempo se ha de producir una muerte será:

$$\Pr\left[j \text{ muera en el tiempo } T_{[i]} \mid j \in R(T_{[i]}), \text{ tiene que ocurrir una muerte en } T_{[i]}\right] = \frac{\exp\{x'_j\beta\}}{\sum_{h \in R(T_{[i]})} \exp\{x'_h\beta\}}$$

Combinando todas estas probabilidades condicionales para todos los tiempos de fallo tendremos la verosimilitud parcial o de Cox:

$$L_C(\beta) = \prod_{i=1}^K \frac{\exp\{x'_j\beta\}}{\sum_{h \in R(T_{[i]})} \exp\{x'_h\beta\}}$$

Cuando hay coincidencia de muertes en un mismo tiempo de fallo una definición alternativa de la función de verosimilitud para tener en cuenta este hecho es:

$$L_C(\beta) = \prod_{i=1}^K \frac{\prod_{j \in D(T_{[i]})} \exp\{x'_j\beta\}}{\left[\sum_{h \in R(T_{[i]})} \exp\{x'_h\beta\} \right]^{d_i}}$$

Donde $D(T_{[i]})$ son el conjunto de individuos que mueren en el tiempo $T_{[i]}$ y d_i es el número de individuos que forman ese conjunto. Esta aproximación en el libro de Kalbfleish y Prentice (1980) se atribuye a Peto (1972) y a Breslow (1974) y es aceptable cuando el número de individuos que mueren en un mismo tiempo no es excesivamente elevado con respecto al número total de individuos a riesgo.

La ecuación más general de la función de verosimilitud parcial, que contempla la presencia de covariables dependientes del tiempo será:

$$L_C(\beta) = \prod_{i=1}^K \frac{\prod_{j \in D(T_{[i]})} \exp\{x'_j(T_{[i]})\beta\}}{\left[\sum_{h \in R(T_{[i]})} \exp\{x'_h(T_{[i]})\beta\} \right]^{d_i}}$$

Es decir, hay que considerar el estado de las covariables específicamente para cada uno de los tiempos de muerte.

Aunque la verosimilitud parcial no es una verdadera función de verosimilitud, en el sentido de que no es proporcional a ninguna función de densidad condicional o marginal de ninguna variable, las estimaciones máximo verosímiles parciales conservarán todas las propiedades asintóticas de cualquier otra estima máximo verosímil. Es decir asintóticamente las estimas se distribuirán como una variable Normal con media el verdadero valor de los parámetros y matriz de varianzas-covarianzas la inversa de menos la matriz de segundas derivadas evaluada en las estimas máximo verosímiles parciales (Kalbfleisch y Prentice, 1980).

En el curso de Ducrocq (2001b) se resumen tres argumentos para justificar el uso de la verosimilitud parcial para inferir los parámetros en un modelo de riesgos proporcionales. El más intuitivo puede ser que una verosimilitud completa se puede factorizar en una parte que solo dependa del efecto de las covariables y otra que dependa del efecto de las covariables y de la función de riesgo base. La primera parte será la verosimilitud parcial de Cox mientras que la segunda aportará poca información sobre el efecto de las covariables, que es el foco de interés en los problemas de regresión, por lo que se puede obviar. Otra justificación, similar a la anterior, fue propuesta de manera más formal por Breslow (1974) quien definiendo la función de riesgo base como una función de escalera en la que el riesgo se mantiene constante entre dos tiempos de muerte sucesivos (como se verá más adelante esto es un *piecewise constant hazard model*), construye una verosimilitud completa que, igualmente, se puede factorizar en una parte que sólo depende del efecto de las covariables y otra que dependerá del efecto de las covariables y de los parámetros de la función de escalera, la primera parte coincidirá con la verosimilitud de Cox. Otra justificación empírica de que la función de verosimilitud parcial funciona es que los resultados obtenidos bajo un modelo paramétrico, que se sabe es el adecuado, y los obtenidos usando la verosimilitud parcial de Cox son muy similares (Chirinos 2003, citando a Kleinbaum (1996)).

En resumen, al trabajar bajo el modelo de Cox, se asume que el efecto que ejercen las covariables se concentra en los instantes en que se producen muertes.

1.2.4.3.2.- Piecewise constant hazard model

Se basa en la partición finita del eje del tiempo $0 < t_1 < t_2 < \dots < t_i < \dots < t_K$ de tal manera que se tendrán K intervalos $(0, t_1], (t_1, t_2], \dots, (t_{i-1}, t_i], \dots, (t_{K-1}, t_K]$. En el i -ésimo intervalo se asumirá que el riesgo base es constante y será igual a h_i , por lo que en total

se tendrán $K h_i$, que serán junto con el efecto de las covariables los parámetros a estimar.

Con este modelo la estimación se puede llevar a cabo mediante máxima verosimilitud. La construcción de la verosimilitud se llevará a cabo de la misma manera que para el caso de un modelo totalmente paramétrico, es decir los datos censurados contribuirán con la función de supervivencia y los no censurados con la de densidad.

Un caso particular de este tipo de modelos fue propuesto por Breslow (1974), en el que la partición del eje del tiempo la definen los propios tiempos de muerte por lo que el riesgo entre dos tiempos de muerte consecutivos se asumirá constante. Como se ha comentado anteriormente, a partir de la verosimilitud definida para este modelo se puede llegar a la función de verosimilitud de Cox.

1.2.5.- Modelos de fragilidad

Hasta ahora, las covariables que se han contemplado en todos los modelos de regresión propuestos son de naturaleza fija. En ciertas situaciones puede ser necesaria la inclusión de efectos de naturaleza aleatoria en los modelos y a este tipo de efectos, en el contexto del análisis de supervivencia, se les llama términos de fragilidad y dan lugar a lo que se conoce como modelos de fragilidad.

El término de fragilidad se define como una cantidad aleatoria, no observable, que afecta multiplicativamente al riesgo. Puede ser específica para cada individuo, por lo que extraerá parte de la variación no observada entre individuos y por lo tanto permite la corrección de la posible discrepancia que pueda haber entre la verdadera varianza de las observaciones y la especificada por el modelo. Cuando el término de fragilidad se define para un grupo de individuos describirá las características no observables que comparten ese grupo de individuos y que actúan sobre el riesgo de todos los miembros del grupo de la misma manera.

En el contexto de la mejora genética, cuando se aplique un modelo animal estaremos ante el primer tipo de términos de fragilidad, mientras que si se trata de modelos macho estaremos ante el segundo tipo. En cualquier caso, la naturaleza aleatoria de la fragilidad proporciona un marco estadístico adecuado para tratar los problemas de mejora genética.

El modelo de riesgos proporcionales de fragilidad (sin estratificación), para el individuo i se puede expresar mediante la ecuación:

$$h_i(t | x'_i(t), z_i) = h_0(t) z_i \exp\{x'_i(t)\beta\}$$

Donde z_i es el término de fragilidad (no-negativo). Así, dos individuos afectados por los mismos efectos fijos, y en el mismo punto del tiempo, tendrán diferente riesgo si tienen diferente término de fragilidad; el que tenga mayor término de fragilidad tendrá mayor riesgo de morir y por tanto será un individuo más “frágil”.

Por la naturaleza aleatoria del término de fragilidad hay que asumir una distribución para él. En el contexto de la mejora genética las distribuciones que se han asumido para este término han sido la Gamma y la log-Normal. La distribución Gamma se propuso inicialmente pues el tratamiento matemático de los modelos era relativamente sencillo, la log-Normal se propuso más tarde pero su tratamiento matemático es más complejo, si bien facilita la consideración de las relaciones de parentesco.

El modelo anterior será equivalente a:

$$h_i(t | x'_i(t), z_i) = h_0(t) \exp\{x'_i(t)\beta + u_i\}$$

Donde u_i sigue una distribución log-Gamma si z_i seguía una distribución Gamma, o una Normal si z_i seguía una log-Normal. El hecho de que u_i se distribuya como una Normal es una propiedad que desde el punto de vista de la mejora genética es de una importancia capital, pues esto permite modelar efectos genéticos, con la distribución correcta de acuerdo con el modelo infinitesimal y la consideración de las relaciones de parentesco, tal cómo hemos comentado.

En los programas de evaluación genética para longevidad en vacuno lechero es muy común que el término de fragilidad tenga dos componentes, una que se distribuye como una Gamma, para ciertos efectos ambientales (rebaño-año-estación) y otra log-Normal, para el efecto genético.

Al igual que para el caso del modelo de efectos fijos los modelos de fragilidad tienen un modelo log-lineal equivalente. Asumamos el modelo de fragilidad con covariables dependientes del tiempo siguiente:

$$h_i(t | x'_{i,1}(t), x'_{i,2}(t), u_i) = h_0(t) \exp\{x'_{i,1}(t)\beta_1 + x'_{i,2}(t)\beta_2 + u_i\}$$

Donde el subíndice 1 hace referencia a las covariables dependientes del tiempo y el subíndice 2 a las no dependientes del tiempo.

Al igual que se indicó antes para el caso de modelos de efectos fijos sin covariables dependientes del tiempo, para el modelo de fragilidad anterior se tendrá que $S_i(t_i | x'_{i,1}(t), x'_{i,2}, u_i) = \exp\{-H_i(t_i | x'_{i,1}(t), x'_{i,2}, u_i)\}$ (siendo t_i la realización para el individuo i de la variable tiempo de fallo) se distribuye de manera uniforme en el intervalo $[0,1]$, y el riesgo acumulado $y_i = H_i(t_i | x'_{i,1}(t), x'_{i,2}, u_i)$ lo hace de manera exponencial con parámetro 1. Siendo el riesgo acumulado:

$$\begin{aligned} y_i = H_i(t_i | x'_{i,1}(t), x'_{i,2}, u_i) &= \int_0^{t_i} h_0(s) \exp\{x'_{i,1}(s)\beta_1 + x'_{i,2}\beta_2 + u_i\} ds \\ &= \exp\{x'_{i,2}\beta_2 + u_i\} \int_0^{t_i} h_0(s) \exp\{x'_{i,1}(s)\beta_1\} ds \end{aligned}$$

Si llamamos $g_i(t_i)$ a $\int_0^{t_i} h_0(s) \exp\{x'_{i,1}(s)\beta_1\} ds$,

$$y_i = H_i(t_i | x'_{i,1}(t), x'_{i,2}, z_i) = \exp\{x'_{i,2}\beta_2 + u_i\} g_i(t_i)$$

El logaritmo de y_i , al igual que para el caso de modelos de efectos fijos sin covariables dependientes del tiempo, se distribuirá como una variable de valor extremo que podemos llamar ε_i .

Por lo que

$$\varepsilon_i = \log(y_i) = x'_{i,2}\beta_2 + u_i + \log(g_i(t_i))$$

Reordenando términos tendremos: $\log(g_i(t_i)) = -x'_{i,2}\beta_2 - u_i + \varepsilon_i$.

Que es un modelo lineal en la escala del logaritmo de $g_i(t)$. Por lo tanto cuando hay covariables dependientes del tiempo, la escala (el modo de acumular riesgo atribuible a la función de riesgo base y a las covariables dependientes del tiempo) es específica para cada individuo o para cada grupo de individuos con las mismas covariables dependientes del tiempo.

La esperanza de la variable $\log(g_i(t_i))$ será $E[\log(g_i(t_i))] = -x'_{i,2}\beta_2 - E(u_i) + \gamma_E$ ($\gamma_E =$ - cte. de Euler) y su varianza $Var[\log(g_i(t_i))] = \sigma_u^2 + \frac{\pi^2}{6}$.

Un caso particular de este modelo es el modelo de Weibull sin covariables dependientes del tiempo. En este caso el modelo log-lineal anterior será:

$$\log(H_0(t_i)) = -x'_{i,2}\beta_2 - u_i + \varepsilon_i$$

Y como el riesgo acumulado bajo un modelo de Weibull es $H_0(t_i) = t_i^\rho \exp\{\rho \log(\lambda)\}$

$$\log(t_i) = -\log(\lambda) - \frac{1}{\rho} x'_{i,2}\beta_2 - \frac{1}{\rho} u_i + \frac{1}{\rho} \varepsilon_i$$

$$\text{Con } E[\log(t_i)] = -\log(\lambda) - \frac{1}{\rho} x'_{i,2}\beta_2 - \frac{1}{\rho} E(u_i) + \frac{1}{\rho} \gamma_E \quad \text{y} \quad \text{Var}[\log(t_i)] = \frac{1}{\rho^2} \left(\sigma_u^2 + \frac{\pi^2}{6} \right)$$

Es decir, el modelo de Weibull sin covariables dependientes del tiempo es un modelo log-lineal del tiempo de fallo. Cuando haya covariables dependientes del tiempo ya no será así.

1.2.5.1.- Estimación bajo modelos de fragilidad

Para abordar la estimación de los componentes de varianza de los términos de fragilidad y, en algunos casos, predecir también sus efectos hay aproximaciones tanto frecuentistas como bayesianas.

Hay dos aproximaciones frecuentistas. Una hace uso del algoritmo EM, para de manera iterativa predecir los efectos de los términos de fragilidad y estimar los efectos de las covariables y de los parámetros de la función de riesgo base (Klein, 1992); la otra aproximación se basa en la integración de los términos de fragilidad en la función de verosimilitud, por lo tanto esta segunda aproximación no permite predecir los efectos de los términos de fragilidad.

Las aproximaciones usadas dentro del campo de la mejora genética son bayesianas. El Survival Kit (Ducrocq y Sölkner, 1994) implementa las técnicas descritas por Ducrocq y Casella (1996). En este programa, en función del tipo de *prior* que se asuma para el término de fragilidad y del modelo que se trate (Weibull o Cox) se lleva a cabo la estimación de componentes de varianza de una manera o de otra. Así, cuando se usa un modelo Weibull y *a priori* se asume una distribución Gamma para el término de fragilidad, el efecto de los términos de fragilidad se puede integrar algebraicamente en la función de densidad posterior conjunta y las estimas del efecto de las covariables, de los parámetros de la función de riesgo base y del parámetro de dispersión de la

distribución Gamma se obtienen calculando el máximo de esta distribución (máximo *a posteriori*). Si con el modelo de Weibull *a priori* se asume una distribución log-Normal para el término de fragilidad, la marginalización algebraica no se puede llevar a cabo, pero se puede llegar a una aproximación de la distribución posterior marginal del parámetro de dispersión del término de fragilidad, usando la técnica de integración Laplaciana. Una vez estimado el parámetro de dispersión (generalmente como la moda de su distribución marginal), la predicción de los efectos del término de fragilidad y la estimación del efecto de las covariables y de los parámetros de la función de riesgo base, se lleva a cabo maximizando la distribución posterior conjunta de todos ellos, condicional a la estima previa del parámetro de dispersión (método bayesiano empírico). Cuando el modelo que se emplea es el de Cox, la integración algebraica no se podrá llevar a cabo, por lo que tanto para un *prior* Gamma como para un *prior* log-Normal la estimación del parámetro de dispersión se hará usando la integración Laplaciana como método aproximado para llegar a la moda de la distribución posterior marginal del mismo, y una vez estimado este parámetro la predicción de los efectos de los términos de fragilidad y la estimación del efecto de las covariables se hace maximizando su distribución posterior conjunta condicional a la estima del parámetro de dispersión.

Como se ha visto, cuando se asume un *prior* log-Normal para el término de fragilidad, que es el que se requiere para trabajar bajo las premisas del modelo infinitesimal, la integración Laplaciana juega un papel fundamental en la estimación de componentes de varianza bajo la aproximación del Survival Kit. Existe un problema con la implementación actual de la integración Laplaciana en el Survival Kit y es que no permite estimar los componentes de varianza para varios efectos aleatorios a la vez. Como se discutirá en el siguiente punto, puede ser interesante, especialmente en modelos macho, que el término de fragilidad tenga dos componentes, uno para modelar el efecto de macho y otro que recoja de manera explícita la parte de variación genética total no contemplada por el efecto de macho ($\frac{3}{4}$ de la varianza genética aditiva total), además de otros efectos aleatorios de naturaleza no genética aditiva que afectan específicamente a cada individuo con dato.

La otra aproximación Bayesiana, para hacer inferencia con los modelos de fragilidad, se basa en métodos de Monte Carlo (Korsgaard y col., 1998) y no hay limitación en el número de efectos aleatorios para los que estimar los componentes de

varianza. Su principal inconveniente es que tiene mayores necesidades de cómputo, lo que hace que en situaciones donde el volumen de datos es muy grande no sea aplicable.

1.2.5.2.- ¿Es necesario incluir un residuo en el término de fragilidad?

Para la discusión de este punto asumiremos un modelo Weibull sin variables dependientes del tiempo y sin estratificación. Como se ha visto antes, este modelo es equivalente a un modelo lineal del logaritmo del tiempo. Consideremos que el término de fragilidad en este modelo solo incluye el efecto de macho (s). Por lo tanto el modelo log-lineal para el dato del individuo i hijo del macho j será:

$$\log(t_i) = -\log(\lambda) - \frac{1}{\rho} x'_{i,2} \beta_2 - \frac{1}{\rho} s_j + \frac{1}{\rho} \varepsilon_i$$

Donde s_j se distribuye de manera Normal con media cero y matriz de varianzas-covarianzas $A\sigma_s^2$, siendo A la matriz del doble de las relaciones de parentesco entre los machos.

Bajo este modelo, Ducrocq y Casella (1996) indican que es el término ε_i (cuya distribución es la de valor extremo) el que incluirá de forma implícita $\frac{3}{4}$ de la varianza genética aditiva total, que es la parte de la variación genética aditiva no explicada por el

efecto de macho. Por lo tanto $Var(\varepsilon_i) = \frac{\pi^2}{6} = 1.645$ deberá ser mayor o igual que tres veces la varianza asociada a macho. Dada la baja heredabilidad de los caracteres de longevidad, la varianza asociada al término de macho generalmente va a ser muy pequeña y por lo tanto es de esperar que esta asunción se cumpla.

Andersen y col. (2000) estudian desde un plano totalmente teórico los modelos de fragilidad aplicados a la mejora genética. En este trabajo se hace un especial hincapié en el hecho de que para que un modelo macho tenga un modelo animal equivalente, acorde con las premisas del modelo infinitesimal, es necesario que se incluya un componente extra en el término de fragilidad, particular para cada animal con dato, y que en el proceso de estimación se imponga la restricción de que la varianza asociada a ese término ha de ser mayor o igual que tres veces la varianza de macho. Se fuerza a que sea ese término el que incluya explícitamente la parte de la variación genética aditiva total no considerada por el efecto de macho; además de otros efectos de

naturaleza ambiental (inobservables) que no se han contemplado en el modelo (heterogeneidad individual no genética aditiva).

La ecuación de este modelo en la escala del logaritmo del tiempo para el caso Weibull y sin covariables dependientes del tiempo ni estratificación será:

$$\log(t_i) = -\log(\lambda) - \frac{1}{\rho} x'_{i,2} \beta_2 - \frac{1}{\rho} s_j - \frac{1}{\rho} e_i + \frac{1}{\rho} \varepsilon_i$$

Donde e_i es el componente extra del término de fragilidad y que se distribuirá de manera Normal con media cero y matriz de varianzas-covarianzas $I\sigma_e^2$.

En trabajos de simulación, Damgaard y col. (2003) han estudiado el efecto de la inclusión de este efecto extra (que se suele llamar residual) en el término de fragilidad. Sus resultados indican que si no se incluye el término residual se producen sesgos en las estimaciones de los parámetros que describen la función de riesgo base, del efecto de las covariables y de la varianza asociada al efecto de macho. En cuanto a la predicción de los efectos de macho no detectan sesgos, al usar el modelo sin término residual, cuando se comparan las predicciones con los valores verdaderos.

A nivel práctico se ha observado un comportamiento inestable de las evaluaciones genéticas de sementales seleccionados en base a una primera tanda de hijas, cuando en un segundo análisis se incorporan los datos de un nuevo lote de hijas en el que la proporción de censura es muy elevada (Ducrocq, 1999a; VanRaden y Powell, 2002). En el trabajo de Damgaard y col. (2003) también se trata de estudiar este efecto. Para ello se hace una evaluación de 100 machos cada uno con 100 hijas, se seleccionan a los 10 mejores en base a un modelo con término residual y a otro sin término residual. A continuación se añaden 900 hijas (con un nivel de censura muy alto) para cada uno de estos 10 toros y se comparan (dentro de modelo) las evaluaciones de estos 10 toros cuando tienen 100 hijas, cuando tienen 1000 y con los valores reales con los que se hace la simulación. Se observa que con el modelo que incluye término residual los valores predichos están más cercanos a los reales, y casi no hay diferencias entre las evaluaciones cuando tienen 100 o cuando tienen 1000 hijas. Si el modelo no incluye el término residual se producen sesgos (hacia arriba) en las evaluaciones, y además las evaluaciones son inestables, en el sentido de que hay cambios en el ranking de los 10 mejores toros cuando se pasa de una situación de 100 hijas a otra de 1000.

En el contexto del modelo animal, como toda la variación genética aditiva va a ser recogida por el efecto asociado al animal, la inclusión del término residual en el

modelo permite corregir por una posible fuente de variación ambiental no contemplada por el resto de factores del modelo.

1.2.5.3.- Definiciones de heredabilidad

La definición de la heredabilidad en los modelos de fragilidad usados para describir datos de longevidad en poblaciones animales no es obvia, pues en la escala original el modelo no es lineal y en las transformaciones lineales no se describe la variable fenotípica observada sino una función de ella. Además en estas transformaciones lineales aparecen términos aleatorios que no siguen una distribución Normal, ε_i .

1.2.5.3.1.- Heredabilidad en la escala de log-frailty

Korsgaard y col. (2002) desarrollan, para el modelo de fragilidad animal log-Normal con término residual, las expresiones teóricas del mejor predictor, de la precisión de la evaluación genética y de la respuesta a la selección basada en el mejor predictor. En todas estas expresiones aparece un término que llaman heredabilidad en la escala del logaritmo del término de fragilidad (log-frailty) o escala con distribución Normal. Esta heredabilidad es el ratio entre la varianza atribuible al efecto aditivo y la varianza total del logaritmo del término de fragilidad.

Asumamos un modelo de fragilidad animal log-Normal Weibull sin covariables dependientes del tiempo y sin estratificación. En este caso el riesgo para el individuo i vendrá dado por:

$$h_i(t | x'_i, z_i) = h_0(t) z_i \exp\{x'_i \beta\}$$

Donde $z_i = \exp\{a_i + e_i\}$ es una variable log-Normal y $a_i \sim N(0, A\sigma_a^2)$ y $e_i \sim N(0, I\sigma_e^2)$, por lo tanto $\log(z_i) = a_i + e_i$. Para este modelo la heredabilidad en la escala de log-

frailty o h_{nor}^2 se define como $\frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$. Si se tratase de un modelo macho sería $\frac{4\sigma_s^2}{\sigma_s^2 + \sigma_e^2}$.

Una ventaja de esta definición es que se puede interpretar como la heredabilidad para un carácter gaussiano: Coeficiente de regresión lineal de una variable sobre el valor aditivo. El inconveniente es que, en este caso, esa variable no es observada, se trata del logaritmo del término de fragilidad. Por lo tanto, a pesar de que la interpretación puede ser equivalente, un valor concreto para esta heredabilidad no implica lo mismo (en el

sentido de respuesta esperada) que ese mismo valor estimado para un carácter gaussiano.

1.2.5.3.2.- Heredabilidad en la escala de log-g(t)

La primera definición de heredabilidad que se dio para caracteres de supervivencia fue un caso particular de ésta, en el trabajo de Ducrocq y Casella (1996), para el modelo macho Weibull sin covariables dependientes del tiempo ni estratificación. Como ya se ha visto, la transformación log-lineal de este modelo es:

$$\log(t_i) = -\log(\lambda) - \frac{1}{\rho} x'_{i,2} \beta_2 - \frac{1}{\rho} s_j + \frac{1}{\rho} \varepsilon_i$$

De esta manera la definición de heredabilidad en la escala del logaritmo del tiempo de muerte será:

$$h^2_{\log(t)} = \frac{4\sigma_s^2 / \rho^2}{\text{Var}[\log(t)]} = \frac{4\sigma_s^2}{\sigma_s^2 + \frac{\pi^2}{6}}$$

Para el caso más general de cualquier modelo macho de riesgos proporcionales de fragilidad con covariables dependientes del tiempo la escala es diferente pues la transformación log-lineal del modelo, como ya se vio, es:

$$\log(g_i(t_i)) = -x'_{i,2} \beta_2 - s_j + \varepsilon_i$$

Donde $g_i(t_i) = \int_0^{t_i} h_0(s) \exp\{x'_{i,1}(s) \beta_1\} ds$

$$h^2_{\log-g(t)} = \frac{4\sigma_s^2 / \rho^2}{\text{Var}[\log(t)]} = \frac{4\sigma_s^2}{\sigma_s^2 + \frac{\pi^2}{6}}$$

Si el modelo incluye término residual la heredabilidad será, $\frac{4\sigma_s^2}{\sigma_s^2 + \sigma_e^2 + \frac{\pi^2}{6}}$.

Esta definición de heredabilidad no se puede interpretar como el coeficiente de regresión de una variable sobre el valor aditivo, como se podía hacer en el caso anterior, ya que aquí estamos ante una variable que no es Normal. Por lo tanto, un valor obtenido usando esta definición de heredabilidad no implica lo mismo (en el sentido de respuesta

esperada) que ese mismo valor estimado para un carácter gaussiano, al igual que ocurría para la heredabilidad en escala *log-frailty*.

1.2.5.3.3.- Heredabilidad en la escala original

En general, el cálculo de las precisiones de las evaluaciones es un requisito fundamental en los esquemas de mejora genética de muchas especies domésticas. Al igual que para caracteres lineales, para caracteres de supervivencia hay métodos para computar la varianza del error de predicción, lo que permite calcular de manera exacta la precisión de la evaluación genética. Para el cálculo de la varianza del error de predicción es necesario la inversión de una matriz que tiene una dimensión de como mínimo el número de individuos que se están evaluando. Esta operación, cuando la cantidad de individuos a evaluar es muy grande, implica tiempos de cómputo prohibitivos y, por eso, para especies como el vacuno lechero, en las que el número de individuos a evaluar es muy elevado, se recurre al cálculo de las precisiones de la evaluación de manera aproximada, usando fórmulas derivadas de la teoría de índices de selección. Así, es común aproximar la precisión de la evaluación de un toro usando la precisión de una prueba de progenie en el que participan n hijas sólo emparentadas por el lado paterno. Esta precisión tiene como ecuación:

$$R = \frac{nh^2}{(n-1)h^2 + 4}$$

En esta fórmula aparece la heredabilidad del carácter y n , que en el contexto del análisis de supervivencia sería el número de hijas no censuradas del toro para el que se está calculando la precisión.

Cuando en la fórmula del cálculo aproximado de la precisión se emplea la heredabilidad en la escala del logaritmo de $g_i(t_i)$ las precisiones así calculadas no se corresponden con las verdaderas obtenidas cuando se computan usando la varianza del error de predicción (Ducrocq, 1999a). Por lo tanto se plantea una definición de la heredabilidad que funcione de manera correcta para hacer esta aproximación. Inicialmente, para resolver este problema se propuso lo que se llamó heredabilidad en escala original (Ducrocq, 1999a), que se desarrolló para el modelo macho Weibull sin covariables dependientes del tiempo ni estratificación. Como ya se ha visto la transformación log-lineal de este modelo es:

$$\log(t_i) = -\log(\lambda) - \frac{1}{\rho} x'_{i,2} \beta_2 - \frac{1}{\rho} s_j + \frac{1}{\rho} \varepsilon_i$$

La transformación de la escala logarítmica a la escala observada de tiempo será:

$$t_i = \exp \left\{ -\log(\lambda) - \frac{1}{\rho} x'_{i,2} \beta_2 - \frac{1}{\rho} s_j + \frac{1}{\rho} \varepsilon_i \right\}$$

$$t_i = \exp \left\{ -\log(\lambda) - \frac{1}{\rho} x'_{i,2} \beta_2 \right\} \exp \left\{ -\frac{1}{\rho} s_j + \frac{1}{\rho} \varepsilon_i \right\}$$

$$t_i = K_i \exp \left\{ -\frac{1}{\rho} s_j + \frac{1}{\rho} \varepsilon_i \right\}$$

De esta manera la heredabilidad en la escala original será, $h_o^2 = \frac{4A}{B}$, siendo

$$A = \text{Var} \left[K_i \exp \left\{ -\frac{1}{\rho} s_j \right\} \right]$$

$$B = \text{Var} \left[K_i \exp \left\{ -\frac{1}{\rho} s_j + \frac{1}{\rho} \varepsilon_i \right\} \right]$$

Las K_i saldrán al cuadrado del operador varianza y en el cociente se simplificarán. Para el cálculo de las varianzas de las funciones del efecto macho y del efecto macho más el término que se distribuye como una variable de valor extremo se recurre al “método delta” de tal manera que tendremos que la heredabilidad en escala original es:

$$h_o^2 = \frac{4\sigma_s^2}{\left[\exp \left\{ \frac{1}{\rho} \gamma_E \right\} \right]^2 \left[\sigma_s^2 + \frac{\pi^2}{6} \right]} = \left[\exp \left\{ \frac{1}{\rho} \gamma_E \right\} \right]^{-2} h_{\log(t)}^2$$

Por lo tanto estamos ante una ecuación para la heredabilidad que es función de la heredabilidad en escala del logaritmo del tiempo, de uno de los parámetros de la función de riesgo base y de menos la constante de Euler (γ_E).

Esta ecuación sólo será una aproximación a la heredabilidad en escala original cuando se use el modelo para el cual se ha desarrollado (Weibull sin covariables dependientes del tiempo ni estratificación). Cuando el modelo no sea éste, su interpretación como heredabilidad en escala original no va a ser adecuada, como señalan Korsgaard y col. (1999). A pesar de esto, se ha usado en modelos con covariables dependientes del tiempo y las precisiones obtenidas de manera aproximada

están muy próximas a las verdaderas. Ducrocq (1999a) señala que el uso de esta heredabilidad con el objetivo de aproximar las precisiones de las evaluaciones tiene que ser comprobado dentro de cada población en la que se estudie la longevidad pues observan que cuando por ejemplo $\rho < 1$ esta heredabilidad no funciona bien para aproximar las precisiones de las evaluaciones.

A pesar de que la heredabilidad en escala original parece ser que se comporta adecuadamente en el cálculo aproximado de las precisiones de la evaluación genética, su interpretación no es equivalente a la heredabilidad de un carácter gaussiano pues está definida para un carácter que, además de no distribirse de manera Normal, se explica a través de un modelo multiplicativo.

1.2.5.3.4.- Heredabilidad efectiva

El hecho de que la heredabilidad en escala original dependa sólo de uno de los parámetros de la distribución de Weibull no es una propiedad deseable, pues dada la relativamente alta correlación negativa entre los dos parámetros de la función de Weibull, a través de diferentes combinaciones de ellos se pueden obtener los mismos ajustes de los datos, pero dando lugar a diferentes heredabilidades en escala original (Yazdi y col., 2002).

Yazdi y col. (2002) proponen una definición de la heredabilidad alternativa, que llaman heredabilidad efectiva. Para el desarrollo de la fórmula de la heredabilidad efectiva proponen igualar la precisión verdadera, obtenida a partir de la varianza del error de predicción (que obtienen de manera aproximada) de un modelo macho Weibull sin estratificación ni variables dependientes del tiempo, a la precisión calculada mediante la teoría de índices de selección de una prueba de progenie. La ecuación de esta heredabilidad es:

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

Con esta definición de heredabilidad, la fórmula para calcular la precisión de la evaluación genética asumiendo un test de progenie será:

$$R = \frac{n_{unc}}{n_{unc} + \frac{1}{\sigma_s^2}}$$

que depende del número de hijas no censuradas (n_{unc}) del macho para el que se está calculando la precisión.

Puede ser de interés calcular la precisión con la quedará evaluado un animal a medida que el número de hijas que mueren va aumentando. Por eso, una escritura alternativa de la precisión que dependa del número total hijas y de la proporción, p , de hijas que hasta ese momento han muerto es:

$$R = \frac{np}{np + \frac{1}{\sigma_s^2}} = \frac{n}{n + \frac{1}{p\sigma_s^2}}$$

Esto es un caso particular de la precisión de la evaluación por un test de progenie que tiene la forma general $\frac{n}{n+k}$ siendo k una función de la heredabilidad de la forma

$\frac{4-h^2}{h^2}$. Por lo tanto si se iguala el término $\frac{1}{p\sigma_s^2}$ a $\frac{4-h^2}{h^2}$, obtendremos otra definición

de la heredabilidad, que es:

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

A esta definición de la heredabilidad se le llama heredabilidad equivalente h_{equ}^2 . El sentido del término equivalente es que la prueba de progenie para un macho que tiene n hijas daría la misma precisión que para un carácter lineal con heredabilidad igual a la h_{equ}^2 . Como la proporción de datos completos, p , cambia con el tiempo h_{equ}^2 tenderá a su máximo que es h_{eff}^2 , cuando todos los individuos tengan el dato completo.

Al igual que ocurría con la heredabilidad en escala original, tanto la heredabilidad efectiva como la heredabilidad equivalente son parámetros derivados bajo un modelo determinado (Weibull sin estratificación ni covariables dependientes del tiempo) y para una utilidad determinada (aproximar las precisiones de la evaluación genética). Usar estas definiciones de heredabilidad cuando se han usado otros modelos no parece muy adecuado, aunque al igual que para el caso de la heredabilidad en escala original, las precisiones de las evaluaciones que la heredabilidad efectiva permite aproximar, en modelos con covariables dependientes del tiempo, son cercanas a las verdaderas obtenidas computando exactamente la varianza del error de predicción.

1.2.6.- Análisis conjunto de caracteres de supervivencia y caracteres gaussianos

En general los análisis multicarácter son deseables en el contexto de la mejora genética por varias razones como puntualiza Mrode (1996). En muchos programas de mejora genética las decisiones de selección se basan en los resultados de evaluaciones para varios caracteres, que correctamente ponderadas se pueden incluir en un índice. Si los caracteres están correlacionados genética y ambientalmente es conveniente plantear un modelo conjunto que explique los caracteres de interés y que tenga en consideración las relaciones entre ellos. De esta manera se aprovecha la información que unos caracteres tienen sobre otros y al hacer evaluaciones mediante modelos multicarácter las precisiones de las evaluaciones genéticas aumentan. Por otra parte, cuando una población está sometida a selección para algún carácter si se quieren obtener resultados insesgados para las evaluaciones de cualquier otro, es necesario plantear un análisis multicarácter en el que se incluya el carácter criterio de selección.

A pesar de las ventajas del análisis multicarácter con respecto a las del análisis unicarácter, el hacer análisis multicarácter tiene dos inconvenientes serios. En primer lugar es necesario conocer con precisión la estructura de varianzas-covarianzas entre los caracteres, es decir hay que conocer además de las heredabilidades de cada uno de los caracteres, las correlaciones genéticas y fenotípicas entre todos los caracteres que intervienen. Para caracteres lineales en general, y para caracteres lineales y de supervivencia en particular, la estimación precisa de las correlaciones genética y ambiental es en algunas situaciones difícil. En segundo lugar los tiempos de cómputo en análisis multicarácter aumentan con el cubo del número de caracteres implicados en el análisis.

Para el caso que aquí nos atañe, que es el del análisis conjunto de caracteres de supervivencia y caracteres lineales, Ducrocq (1999b) señala cuales son las dificultades conceptuales de este problema. Si para el carácter de longevidad se asume un modelo animal Weibull sin variables dependientes del tiempo, sin estratificación y sin término residual en *log-frailty* hemos visto que la transformación log-lineal de este modelo es:

$$y_{1,i} = \log(t_i) = -\log(\lambda) - \frac{1}{\rho} x_i' \beta_1 - \frac{1}{\rho} a_{i,1} + \frac{1}{\rho} \varepsilon_i$$

Donde, si recordamos, ε_i sigue una distribución de valor extremo. Podemos llamar $\theta_1 = \{\lambda, \rho, \beta_1, a_1\}$ al conjunto de parámetros del modelo que explican los datos de longevidad.

Para el otro carácter que se puede explicar mediante un modelo lineal: $y_{2,i} = x_i' \beta_2 + a_{i,2} + e_i$, donde, e_i sigue una distribución normal, al conjunto de parámetros que describen los datos se le puede llamar $\theta_2 = \{\beta_2, a_2\}$.

La matriz de varianzas-covarianzas en este modelo bicarácter la podemos llamar Σ .

En un contexto Bayesiano de estimación de todos los parámetros se hará a partir de la función de densidad posterior conjunta, que será igual a:

$$p(\theta_1, \theta_2, \Sigma | y_1, y_2) \propto L(y_1, y_2 | \theta_1, \theta_2, \Sigma) p(\theta_1, \theta_2 | \Sigma) p(\Sigma)$$

El problema bajo este enfoque radica en cómo construir la verosimilitud $L(y_1, y_2 | \theta_1, \theta_2, \Sigma)$, pues sería necesaria una distribución bivalente cuyas marginales siguiesen una distribución Normal y una distribución de valor extremo.

Ducrocq (1999b) señala dos posibles situaciones en las que el cálculo de esta función de verosimilitud es posible. Una es un caso particular, en el que ambos caracteres están ambientalmente incorrelacionados, en esta situación la verosimilitud se puede factorizar de la siguiente manera $L(y_1, y_2 | \theta_1, \theta_2, \Sigma) = L(y_1 | \theta_1, \Sigma) L(y_2 | \theta_2, \Sigma)$. La otra situación es el caso general, en el que los caracteres están correlacionados tanto genética como ambientalmente. La solución que se propone consiste en dividir el término residual del carácter de supervivencia en dos componentes una correlacionada con el término residual del carácter lineal y otra que no lo esté; cuando se hace esto, la primera componente se distribuirá siguiendo un modelo Normal y la segunda componente lo hará de acuerdo con una distribución de valor extremo. De acuerdo con esta aproximación la verosimilitud se podrá descomponer de la siguiente manera:

$$L(y_1, y_2 | \theta_1, \theta_2, \Sigma) = L(y_1 | \theta_1, \theta_2, y_2, \Sigma) L(y_2 | \theta_2, \Sigma)$$

Para ambas situaciones Ducrocq (1999b) propone dos estrategias de estimación, la primera cuando la matriz de varianzas-covarianzas es conocida, las estimas de θ_1, θ_2 se puede obtener como la moda de su distribución posterior conjunta condicionada a los componentes de varianza Σ . La otra se centra en la estimación Σ para lo que propone

usar la integración Laplaciana para llegar a la distribución marginal de la matriz de varianzas covarianzas.

A pesar de las dificultades intrínsecas del análisis conjunto de caracteres que se describen a través de modelos de supervivencia y de otros caracteres que se describen mediante modelos lineales, se han propuesto varias alternativas para abordar este análisis multicarácter

1.2.6.1.- Transformación de los datos de longevidad para poder explicarlos a través de las ecuaciones del modelo mixto

Esta estrategia consiste en usar las estimaciones y predicciones de los efectos procedentes de los análisis supervivencia, para con ellas hacer una transformación de los datos originales de longevidad, de tal manera que se puedan explicar mediante un modelo lineal mixto. Los datos transformados, junto con un valor de ponderación asociado a cada uno de ellos, se pueden usar en un análisis multicarácter lineal con otros caracteres de interés igualmente transformados.

El procedimiento de transformación de los datos de longevidad que se ajustan mediante un modelo de Weibull sin covariables dependientes del tiempo fue desarrollado por Ducrocq (2001a). El objetivo de esta transformación es poder explicar los datos transformados mediante un modelo animal del tipo:

$$y_i^* = \mu + a_i + e_i$$

Con varianzas $Var[a] = A\sigma_a^2$ y $Var[e] = diag\left\{\frac{1}{\omega_i}\right\}\sigma_e^2$

El procedimiento para calcular los datos transformados (y_i^*) y sus pesos (ω_i) parte de la ecuación con la que se obtiene el máximo de la distribución posterior conjunta, condicional a la varianza aditiva para el valor aditivo del individuo i .

$$y_i^{\hat{p}} \exp\{x_i'\hat{\beta}\} \exp\{\hat{a}_i\} + \frac{1}{\sigma_a^2} (A^{-1}\hat{a})_i = \delta_i$$

Si expresamos $\exp\{\hat{a}_i\}$ como $\hat{a}_i + \exp\{\hat{a}_i\} - \hat{a}_i$ y llamamos ω_i a $y_i^{\hat{p}} \exp\{x_i'\hat{\beta}\}$ la ecuación anterior se puede escribir como:

$$\omega_i \hat{a}_i + \frac{1}{\sigma_a^2} (A^{-1}\hat{a})_i = \omega_i \left(\frac{\delta_i}{\omega_i} - \exp\{\hat{a}_i\} + \hat{a}_i \right)$$

Que se puede reconocer como la parte de la ecuación del modelo mixto para el valor aditivo del individuo i , del modelo lineal presentado anteriormente, para cuando σ_e^2 es

$$\text{igual a 1 y } y_i^* = \frac{\delta_i}{\omega_i} - \exp\{\hat{a}_i\} + \hat{a}_i .$$

Siendo ω_i el valor de ponderación asociado con el dato del animal i .

Usando esta técnica Besbes y col. (2002) hacen una evaluación conjunta en gallinas de puesta para los caracteres de producción, de la calidad del plumaje y de longevidad.

Ducrocq y col. (2003) propone una modificación de esta transformación en la que el peso asociado al dato de cada individuo se establece a partir de la parte procedente de los datos del elemento de la diagonal de la matriz de información, de esta manera el peso asociado a cada individuo refleja directamente la información que ese dato contiene. Así, las ecuaciones del dato transformado y el peso para ese dato transformado para el individuo i variarán ligeramente.

Recientemente, Tarres y col. (2005) han descrito como abordar la estimación de la correlación genética y ambiental cuando los datos se someten a las transformaciones descritas por Ducrocq y col. (2003). En concreto ellos proponen multiplicar ambos términos del modelo lineal por la raíz cuadrada del valor de ponderación (ω_i), de tal manera que se llega a un modelo en el que no hay heterogeneidad de varianzas residuales. Este trabajo de Tarres y col. (2005) es un estudio de simulación en el que se observa un comportamiento en general bastante aceptable de este método aproximado del cálculo de las correlaciones genéticas y ambientales entre un carácter de supervivencia y otro Normal.

1.2.6.2.- Métodos que incluyen un residuo en el término de fragilidad del modelo de supervivencia

Damgaard (2004) propone un método Bayesiano, resuelto mediante técnicas MCMC, para abordar el análisis conjunto de caracteres lineales y caracteres de supervivencia, y de caracteres umbral y caracteres de supervivencia. Los modelos de supervivencia con los que trabaja son modelos en los que se incluye un término residual en el término de fragilidad. Por lo tanto, esta aproximación tiene cierta similitud con la aproximación propuesta por Ducrocq (1999b) para cuando las correlaciones ambientales no son cero. Es decir los residuos del modelo de supervivencia se descomponen en una

parte que no está correlacionada con los otros caracteres (el término que se distribuye de acuerdo con una distribución de valor extremo) y otra que si lo está y para la que se asume una distribución Normal. En el trabajo de Damgaard (2004) para el análisis conjunto de un modelo lineal con otro de supervivencia, cuando se construye la función de verosimilitud condicional, se asume que el logaritmo del término de fragilidad del modelo de supervivencia (que incluirá parte del residuo y el efecto genético) y la medida fenotípica del carácter lineal se distribuyen conjuntamente como una Normal bivalente.

1.3.- Revisión bibliográfica de la longevidad por especies

A continuación se hace una revisión del carácter por especies. Empezaremos por el ganado vacuno lechero, que es la especie para la que originalmente se han dado las diferentes definiciones que del carácter existen y en la que, como hemos visto, se han propuesto las diferentes metodologías estadísticas para su estudio; definiciones y metodologías que posteriormente se han extendido a otras especies ganaderas: vacuno de carne, pequeños rumiantes, cerdos y conejos. En esta revisión también se incluye un apartado dedicado a las especies de laboratorio y no se incluye ninguno dedicado a aves pues en general en ellas son muy pocos los trabajos publicados.

1.3.1.- Vacuno Lechero

1.3.1.1.- Importancia económica

Rendel y Robertson (1950) puntualizaron las vías a través de las que un incremento de la longevidad repercute favorablemente sobre la función de beneficios de una explotación: i) reduciendo los costes de reposición por vaca en el rebaño, ii) incrementando la producción media del rebaño a través de un aumento de la proporción de vacas en los grupos de edad más productivos, iii) reduciendo la proporción de novillas que hay que criar y por consiguiente aumentando la proporción de vacas en producción dentro del rebaño, dada la superficie de la explotación y iv) aumentando la proporción de vacas eliminadas por bajo nivel de producción. Las dos últimas vías son mutuamente excluyentes, y es una decisión del ganadero el optar por reducir la tasa de reposición o mantenerla. A pesar de estas puntualizaciones los autores concluyen que en esa época (año 1950) la importancia económica de la longevidad en comparación con la producción es muy pequeña, pues un aumento de la vida media de un año sería equiparable a un aumento de la producción de leche total de 35 galones (159 litros).

Actualmente, dado que la mayor parte de los mercados están sometidos a cuotas de producción, los caracteres funcionales en general y la longevidad, como uno de ellos, en particular, adquieren un mayor peso económico (Essl, 1998).

En la revisión de Essl (1998) se resume que el peso económico de la vida productiva de la vaca (expresado en desviaciones típicas aditivas) comparado con el de la producción de leche (que se fijó en 1) varía según las estimaciones entre 0.25 y 1.01 con un promedio de 0.5.

A pesar de la importancia económica que la longevidad directamente pueda tener como carácter funcional, Strandberg y Sölkner (1996) señalan que la consideración de la longevidad funcional en los programas de mejora genética en vacuno lechero es una vía de mejorar de manera indirecta la salud, la reproducción y otras causas de eliminación involuntaria. Así, es de esperar que repercuta de manera positiva en el objetivo global, que es aumentar los beneficios de la producción.

1.3.1.2.- Longevidad verdadera y funcional

Una revisión en la que se tratan diferentes aspectos de la longevidad en vacuno lechero es la de Essl (1998). Uno de los puntos tratados es el antagonismo entre madurez temprana y longevidad que se ha comprobado experimentalmente que existe. En esta revisión además se propone dividir la vida de los animales en dos periodos. El primero es el periodo en que se incurre en costes sin tener beneficios y va desde el nacimiento hasta el primer parto y el segundo es el periodo productivo, desde este primer parto hasta la eliminación o muerte del animal. Las razones para distinguir entre ambos periodos son dos, por un lado los criterios de eliminación son diferentes en cada uno de los dos periodos, y por otro no suele haber registros de eliminación antes del primer control lechero. Es a la duración del segundo periodo a la que se le llama longevidad, aunque existen otros términos de significado similar como: “duración de la vida productiva”, “vida en el rebaño” o “número de partos”. Otro término relacionado con la longevidad es *stayability*, que como ya se ha comentado fue originalmente definido por Everett (1976 a,b) y se refiere a la capacidad de los animales para permanecer en el rebaño hasta una determinada edad.

Ducrocq (1987) propone diferenciar entre lo que sería “longevidad verdadera” y “longevidad funcional”, basándose en la distinción entre eliminación voluntaria e involuntaria. La longevidad verdadera es la capacidad del animal para evitar tanto la eliminación voluntaria como la eliminación involuntaria, mientras que la longevidad funcional es la capacidad de evitar la eliminación involuntaria. La distinción entre causas voluntarias o involuntarias de eliminación puede ser en algunos casos un tanto subjetiva, aunque en otras situaciones es clara. Por ejemplo, la eliminación por cuestiones meramente productivas siempre será voluntaria, mientras que la eliminación por accidentes o muerte siempre será involuntaria. Por el contrario la eliminación por baja fertilidad puede clasificarse como voluntaria si el animal no es muy productivo o puede ser involuntaria si se trata de un animal muy productivo. Ambos caracteres

(longevidad verdadera y longevidad funcional) deberían ser estudiados pues el primero representa una medida general del valor económico global del animal, aunque es el segundo el que mayor interés tiene desde el punto de vista de la mejora genética, pues mide la capacidad del animal para evitar la eliminación por cuestiones relacionadas con la salud, cuya dependencia de factores externos (políticas de eliminación voluntaria) es menor.

El problema de esta diferenciación radica en que no hay una medida directa de la longevidad funcional pues siempre habrá eliminación voluntaria (por producción) dentro del rebaño (Essl, 1998). La estrategia empleada para llegar a una medida de la longevidad funcional es corregir la longevidad verdadera por el nivel de producción. Como la eliminación por producción es una decisión que depende del rebaño en el que nos encontremos, la forma de corregir será a través de la desviación de cada vaca con respecto a la media de su rebaño. Esta corrección, por simplificación, se suele hacer de manera lineal, aunque hay evidencias de que la relación pudiera no ser lineal (Dekkers y Jairath, 1994, citando a varios autores).

Actualmente parece existir un consenso a la hora de estudiar la longevidad en ganado vacuno lechero, para hacerlo sobre la longevidad funcional (longevidad corregida por producción dentro de rebaño). Esta decisión se justifica debido a que los sesgos en los que se incurren al estimar los parámetros genéticos de este carácter (heredabilidad y correlación genética con producción) son menores que los que se tienen cuando se estudia la longevidad verdadera, en la que siempre existe el efecto directo de la eliminación por cuestiones productivas. No obstante, las estimaciones seguirán siendo sesgadas, pues al corregir por producción se arrastra también cualquier correlación fenotípica que pudiera haber entre producción y longevidad funcional (Dekkers, 1993).

1.3.1.3.- Parámetros genéticos y selección

En cuanto a la estima de la heredabilidad del carácter es altamente dependiente de la definición del mismo y del método de análisis empleado. Incluso dentro de una misma definición del carácter y de un mismo método de análisis, las estimas de la heredabilidad pueden darse en diferentes escalas, que no son comparables entre si. En el trabajo de Ducrocq (1987) se hace una revisión de las estimas de heredabilidad desde principios de los años 60. Encuentra un rango muy amplio de valores (0.008-0.39), imputable a que los métodos de análisis son muy variados, y muchos de ellos no

permiten corregir por efectos ambientales, y a que las definiciones del carácter también son muy diversas. Una revisión de las estimaciones más recientes obtenidas por metodología de análisis de supervivencia se hace en la tesis de Chirinos (2003). En esta revisión la variación es algo menor (0.078-0.298), debido, por un lado, a que se trata siempre de la misma metodología y parámetro (heredabilidad efectiva) y, por otro, a que en la mayoría de los casos se trata de longevidad definida de la misma manera: número de días desde el primer parto hasta la eliminación o muerte del animal, corrigiendo por nivel de producción dentro de rebaño (longevidad funcional).

Inicialmente, en los esquemas de mejora genética del vacuno lechero la longevidad no se consideraba como un objetivo adecuado por el que hacer selección de manera directa. Los argumentos que se daban eran la baja heredabilidad del carácter, que está correlacionado positivamente con la producción y que se trata de un carácter que se mide tarde en la vida de los animales. Por tratarse de un carácter de expresión tardía su consideración implicaría un importante incremento del intervalo generacional, reduciéndose la respuesta por unidad de tiempo para otros caracteres. Además, como se hacía selección por producción con la que se creía que estaba correlacionada se esperaba respuesta indirecta (Smith 1983, citado por Ducrocq 1987). La correlación genética positiva entre producción de leche en primera lactación y longevidad es un hecho documentado (Strandberg, 1985, citado por Essl, 1998), pero hasta la fecha no se ha obtenido respuesta correlacionada en longevidad a pesar de la intensa selección que ha habido para producción. De hecho se ha observado un descenso en la longitud de la vida productiva (Essl, 1998). Este resultado se explica si se acepta que las estimas de correlación genética pudieran ser sesgadas, como se comentó anteriormente, a consecuencia del efecto directo del nivel de producción sobre la longevidad (Dekkers, 1993). De hecho, en estimaciones posteriores de la correlación genética entre longevidad (verdadera y funcional) y producción, siempre se han obtenido valores superiores con la verdadera, y en muchos casos ligeramente negativos con la funcional (Essl, 1998).

La falta de respuesta correlacionada en longevidad al seleccionar por producción, unido a la importancia que recientemente han adquirido los llamados caracteres funcionales, ha hecho que la longevidad se considere de manera directa en la mayoría de los esquemas de mejora, e incluso, Interbull hace evaluaciones internacionales de sementales para longevidad (<http://www-interbull.slu.se/longevity/framesida-long.htm>).

A pesar de que se hacen evaluaciones genéticas y hay selección por longevidad, al tratarse de un carácter de expresión tardía en la vida del animal (cuando se elimina o muere), hay bastantes trabajos en los que se estudian cómo diferentes caracteres, que se manifiestan a edades tempranas, están relacionados con la longevidad. Entre estos caracteres los más estudiados son los de tipo (p.e. Chirinos, 2003; Larroque y Ducrocq, 2001; Sölkner y Petschina, 1999), aunque también se han considerado otros como la resistencia a ciertas enfermedades (infertilidad: Strandberg y Roth, 1999; mamitis: Jensen y col., 1999). Además muchos países cuando hacen selección por longevidad lo hacen a través de índices en los que se consideran los valores genéticos para longevidad conjuntamente con los valores genéticos de caracteres de tipo (Reino Unido: Vissher y col., 1999; Canadá: Jairath y col., 1998).

En relación con estos caracteres de tipo y resistencia a enfermedades hay que citar un trabajo en el que se han detectado dos QTL para longevidad funcional. Uno de ellos se encuentra en la misma región que otros QTLs detectados previamente y relacionados con la inserción de la ubre y la velocidad de ordeño, y ambos están en las mismas regiones que otros QTLs detectados previamente para recuento de células somáticas (Kühn y col., 2003).

1.3.2.- Vacuno de carne

1.3.2.1.- Importancia económica

Cundiff y col. (1992) consideran la longevidad como uno de los cuatro componentes de la producción acumulada de una vaca, los otros tres son la fertilidad y la supervivencia y capacidad de crecimiento de los terneros.

En general los argumentos para el estudio y tratamiento de la longevidad en vacuno de carne son similares a los expresados en vacuno lechero y los que más adelante comentaremos en otras especies. En concreto, Arthur y col. (1993) justifican su estudio pues aumentarla permite reducir el número de vacas que hay que reponer cada año, aumentar el número de animales disponible para la venta, dividir los costes de mantenimiento de la vaca entre un mayor número de unidades de producto obtenidas y aumentar la proporción de vacas que están en las clases de edad más productivas. Una mayor longevidad de los animales del rebaño implica que los intervalos generacionales se alarguen, por lo que las respuestas por unidad de tiempo para otros caracteres se pueden reducir. Este efecto negativo se puede compensar, pues al tener que reponer un menor número de animales la intensidad de selección que se aplicará será mayor.

Recientemente, Wolfová y col. (2005) han estimado los pesos económicos para la raza Charolais bajo diferentes sistemas de producción. Un resultado general de este trabajo es que la componente genética directa de las características del parto y la longevidad de la vaca son los caracteres de mayor importancia económica bajo todos los sistemas de producción estudiados. Los siguientes caracteres varían en función del sistema de que se trate, así en producción en pureza con venta de terneros al destete los siguientes caracteres en importancia son la fertilidad de las vacas y el peso al destete de los terneros. Si se trata de un sistema de producción en pureza pero los terneros que no se usan para reposición se ceban en la propia explotación, después de la longevidad y las características del parto siguen en importancia económica los caracteres que influyen en todo el sistema de costes (peso adulto de las vacas, peso de los terneros a 365 días de edad) y los caracteres que influyen en el número de terneros que se cebarán y en los ingresos por venta de carne (fertilidad de las vacas, pérdidas de terneros, rendimiento de la canal).

1.3.2.2.- Variabilidad, heterosis y parámetros genéticos

En el trabajo de Arthur y col. (1993) se estudia el efecto del tipo genético (Hereford, Cruzado de razas cárnicas y Cruzado de razas lecheras y cárnicas) sobre la longevidad y la producción a lo largo de la vida de los animales. Se observa que son los animales cruzados de razas lecheras y cárnicas los de mayor longevidad, seguidos de los cruzados de razas cárnicas y por último los Hereford. En este trabajo una explicación de esta variabilidad genética se hace a través de la consideración de los efectos heteróticos que aparecen en los animales cruzados, efectos que sobre la longevidad en ganado vacuno de carne fueron previamente estimados por otros autores (Núñez-Dominguez y col., 1991). El trabajo de Núñez-Dominguez y col. (1991) estudia datos procedentes de un cruce dialélico completo entre las razas Angus, Hereford y Shorthorn. Como resultados más destacados se pueden señalar las diferencias para longevidad (edad a la eliminación medida en años) y supervivencia hasta los doce años (proporción de animales presentes en el rebaño de los que iniciaron el experimento) entre las razas puras; y la superioridad de los cruzados con respecto a las razas puras. Esta superioridad se estimó en 16.7% para la supervivencia y 1.36 años (16.2 % de heterosis) para la longevidad. Con respecto a la medida de longevidad hay que decir que los animales que en el momento del análisis estaban vivos fueron tratados como si hubiesen muerto, por lo que habrá sesgos en estos resultados. Otro trabajo más reciente en el que igualmente

se obtienen estimas positivas para la heterosis en un cruce dialélico es el de Vega-Murillo y col. (1999), en este trabajo la metodología empleada es la del análisis de supervivencia que es más adecuada dadas las características de los datos de longevidad.

En vacuno de carne la mayoría de los trabajos encontrados definen el carácter longevidad como permanencia en el rebaño a una edad o número de partos determinados, es decir, como un carácter de expresión binaria. El trabajo más reciente encontrado en la bibliografía es el de Silva y col. (2003) en la raza Nelore, en el que se obtienen unas heredabilidades para presencia en el rebaño (la vaca tiene un parto) a los 5, 6 y 7 años de edad, condicionado a que la vaca parió a una edad anterior, de 0.12, 0.11 y 0.17. Doyle y col. (2000) en un trabajo en la raza Angus, en el que estudiaban la heredabilidad de la fertilidad en las novillas y su relación genética con la permanencia en el rebaño hasta el quinto parto, estimaron que para este carácter de *stayability* la heredabilidad era de 0.15. En cuanto a su relación genética con la fertilidad de las novillas se estimó que no era lineal, los animales con alto valor genético para fertilidad tuvieron alto valor genético para *stayability*, los que tuvieron un valor genético intermedio para fertilidad tuvieron el peor valor genético para *stayability* y los animales con peor valor genético para fertilidad tuvieron un valor genético intermedio para *stayability*. En otro trabajo más antiguo (Snelling y col., 1995) también se hace una estimación de la heredabilidad de la longevidad en dos poblaciones de ganado vacuno de carne. En este trabajo igualmente se define la longevidad de las vacas como permanencia en el rebaño (*stayability*) hasta diferentes partos (2° = STAY2, 5° = STAY5, 8° = STAY8 y 11° = STAY11), condicionado a que la vaca haya tenido al menos un parto. Se emplearon diferentes modelos (animal o macho) y métodos de análisis que tuvieron en consideración la naturaleza binaria del carácter (Máxima verosimilitud marginal² y Método R). En este trabajo fue el modelo animal resuelto usando el Método R el que dio estimas más elevadas, resultando en valores que variaron entre 0.025 (STAY2) y 0.23 (STAY5) en una población y 0.12 (STAY5) y 0.21 (STAY2) en la otra. Diaz y col. (2002) estimaron la heredabilidad usando métodos de análisis de supervivencia en dos bases de datos de la raza Avileña Negra Ibérica. En este trabajo la longevidad se definió como el tiempo entre el primer y el último parto. Las estimas de heredabilidad en escala original fueron 0.041 y 0.032. Hyde (2000) estimó la

² Es el nombre que Höschele y col. (1987) usan para denotar a un método bayesiano en el que se obtienen como estimas puntuales de los componentes de varianza la moda de su distribución marginal posterior cuando se usan *flat priors*

heredabilidad para la longevidad de la raza Red Angus norteamericana. La longevidad fue definida como la longitud de la vida reproductiva, medida como el tiempo en años desde que una vaca tiene su primer parto hasta que es eliminada por infertilidad o es censurada. El valor estimado en escala logarítmica³ fue de 0.053. La metodología empleada fue una versión del análisis de supervivencia propuesta por Kachman (1999), que no discutiremos con detalle en este trabajo. Pero diremos que se basa en la resolución de manera iterativa de las ecuaciones del modelo mixto, usando como observaciones unas funciones de las estimas de los parámetros (en la iteración actual) del modelo de supervivencia.

1.3.3.- Pequeños rumiantes

En general son pocos los trabajos que de estas especies se han encontrado en la bibliografía. Uno de ellos es un estudio de la *stayability* hasta la 4^a, la 5^a y la 6^a estación de cría en diferentes razas caprinas en México (Pérez-Razo y col., 2004). En este trabajo además de una serie de factores ambientales y de estacionalidad también se estudia el efecto de la raza sobre la supervivencia, encontrándose diferentes probabilidades de supervivencia en función de la raza a la que pertenezcan los animales. En otro trabajo publicado recientemente y realizado en ovejas de producción cárnica (Casas y col., 2004) se estudia el efecto de la raza y de su interacción con el momento de la cubrición dentro de la estación de cría de otoño, para una gran cantidad de caracteres productivos en animales cruzados, entre los que incluye la supervivencia de la oveja. Para este carácter tanto la raza de la madre como la del padre de las ovejas cruzadas fueron efectos significativos en la determinación de la longevidad. Por lo tanto, al igual que en el trabajo anterior en cabras, se puede decir que existe variabilidad genética entre razas para la longevidad en estas especies.

El trabajo más reciente, en estas especies, en el que se aborda la estimación de parámetros genéticos para diferentes definiciones de longevidad es el de El-Saied et al. (2005). Este estudio se lleva a cabo en ovejas de raza Churra. Se trata de una raza de aptitud lechera pero en la que los ingresos por venta de corderos son muy importantes. Sólo se incluyeron datos de animales con registros completos (no censurados) y el método de estimación fue REML. En general las heredabilidades para diferentes

³ En esta revisión cuando hablemos de heredabilidad en escala logarítmica nos estaremos refiriendo al parámetro descrito en el apartado [1.2.5.3.2](#), heredabilidad en la escala de $\log-g(t)$. Excepto para el trabajo de Guo et al. (2001), en el que se aclara específicamente a que nos referimos.

definiciones de longevidad fueron bajas, 0.02 (vida, medida como el número de días desde el nacimiento a la muerte), 0.05 (vida productiva, medida como el número de días desde el primer parto a la muerte) y 0.06 (vida útil, medida como el número de días en lactación). Igualmente se estiman las correlaciones genéticas entre estos caracteres y otros relacionados con la producción, como la producción lechera por día de vida, de vida productiva y de vida útil. Estas correlaciones variaron entre 0.01 (vida productiva – producción lechera por día de vida productiva) y 0.082 (vida útil – producción lechera por día de vida).

1.3.4.- Cerdos

1.3.4.1.- Importancia económica

Rodríguez-Zas y col. (2003) justifican la consideración del carácter debido a que su aumento implicaría: cambios en la estructura de edad del rebaño, teniendo una mayor proporción de animales en las clases de edad más productivas (mayores tamaños de camada y menos días improductivos (Lucia y col., 2000)); aumento de la proporción de animales que han adquirido inmunidad a las enfermedades del rebaño; aumento del valor de las cerdas en el momento de su eliminación y reducción de los costes directos por reposición.

En trabajos en los que se cuantifica el peso económico de la tasa de reposición o de la longevidad (de Vries, 1989; Sehested, 1996; Rodríguez-Zas y col., 2003), como un resultado general se puede decir que su importancia es moderada, aunque ésta depende del nivel de otros caracteres considerados en las funciones de beneficios y de las condiciones de mercado. Sehested (1996) encuentra que el punto de máximo beneficio del momento de la reposición es a los 6.5 partos, si se tiene en consideración el mayor nivel genético de los animales con los que se va a hacer la reposición. Este nivel en la actualidad está lejos de ser alcanzado. Por ejemplo, los resultados del BD-porc para España en el año 2004 (<http://www.bdporc.irta.es/informes/PartPublica/Datos%20públicos%20Año%202004.htm>) muestran que la proporción de hembras que fueron baja fue del 44.6% y el número de camadas destetadas en promedio por cerda fue de 4.02.

Por lo tanto, dado el interés del carácter, existe una amplia bibliografía para estudiar cómo diferentes factores ambientales y de manejo determinan la longevidad de las hembras y caracteres relacionados con ella (p.e. número total de lechones

producidos) (Brandt y col., 1999; Tantasuparuk y col., 2001; Jørgensen y Sørensen, 1998; Jungst y col. 1988; O'Dowd y col., 1997).

1.3.4.2.- Variabilidad genética y selección

Se han encontrado trabajos en los que se hacen estudios comparativos de diferentes razas, líneas y cruces de animales en cuanto a su longevidad y en todos ellos se ha observado un efecto del tipo genético sobre la longevidad (Hall y col., 2002; Rodríguez-Zas y col., 2003).

La estimación de la heredabilidad también se ha abordado en diferentes estudios. Yazdi y col. (2000a) estimaron, usando metodología de análisis de supervivencia, diferentes heredabilidades, en función del modelo empleado, para la población Landrace sueca. Los valores estimados en escala original variaron entre 0.11 y 0.27, y en escala logarítmica entre 0.05 y 0.12, en función del modelo empleado. Para la población Yorkshire sueca Yazdi y col. (2000b) igualmente estimaron la heredabilidad en escala original y logarítmica usando metodología de análisis de supervivencia; los valores variaron entre 0.21 y 0.31 para la primera y 0.09 y 0.14 para la segunda, en función del tipo de modelo usado. Serenius y Stalder (2004) estimaron las heredabilidades para longevidad funcional y verdadera para las poblaciones Landrace y Large White finlandesas. Para la población Landrace la heredabilidad de la longevidad verdadera estimada por métodos de análisis de supervivencia y definida en escala logarítmica fue de 0.16; para la longevidad funcional la heredabilidad definida en esta misma escala fue de 0.17. Para la población Large White las estimas de la heredabilidad en escala logarítmica fueron de 0.17 y 0.19 para longevidad verdadera y funcional respectivamente. Otro trabajo en el que se estima la heredabilidad para una población Landrace en Estados Unidos es el de Guo y col. (2001). En este trabajo la metodología empleada es de modelo lineal mixto pero considerando la información parcial que hay en los datos censurados. Los resultados fueron de 0.25 en la escala original y de 0.22 en la escala logarítmica⁴. López-Serrano y col. (2000) estimaron las heredabilidades para

⁴ Estas definiciones de la heredabilidad son el ratio entre la varianza genética aditiva y la varianza fenotípica, estimadas sin hacer ninguna transformación de los datos o aplicando una transformación logarítmica. El caso logarítmico podría interpretarse como un modelo de riesgos proporcionales en el que la función de riesgo base se describe a través de un modelo log-Normal. Por lo que la heredabilidad en esa escala sería equivalente al parámetro descrito en el apartado [1.2.5.3.2](#), heredabilidad en la escala de log-g(t), sustituyendo el término $\pi^2/6$ por la varianza residual.

la *stayability* hasta el segundo (STAY12) y tercer (STAY123) parto en poblaciones Landrace y Large White alemanas. Las heredabilidades para STAY12 y STAY123 en la población Landrace fueron de 0.07 y 0.11 respectivamente, mientras que esos mismos valores para la población Large White fueron de 0.08 y 0.10. Tholen y col. (1996a) estudiaron la *stayability* en dos rebaños comerciales, uno de ellos formado por animales de las razas Landrace y Large White, y el otro formado por una población sintética entre estas dos mismas razas. En este trabajo se define la *stayability* verdadera como la permanencia o no en el rebaño desde el 1^{er} parto hasta el 2^o (STAY12), 3^o (STAY13) y 4^o (STAY14). Para estos tres caracteres estiman la heredabilidad dentro de cada uno de los rebaños. Los valores estimados variaron entre 0.02 (STAY12) y 0.09 (STAY14) para el rebaño formado por las razas Landrace y Large White y 0.05 (STAY12) y 0.08 (STAY13 y STAY14) para el rebaño sintético. Igualmente, para ambos rebaños definen la *stayability* funcional como la permanencia o no en el rebaño desde el primer parto hasta el tercero (STAY_c13) y el cuarto (STAY_c14), corrigiendo por el número de nacidos vivos promedio desde el primer parto hasta el último (tercero o cuarto). Para estos dos caracteres las estimas fueron de 0.07 (STAY_c13) y 0.09 (STAY_c14) para el rebaño formado por las razas Landrace y Large White y 0.09 (STAY_c13) y 0.06 (STAY_c14) para el rebaño sintético. Con respecto a este trabajo hay que decir que las estimaciones para el primer rebaño deben interpretarse como intermedias a las que corresponderían a las dos razas que lo formaban, pues éstas se mantuvieron independientes.

En cerdos existe la misma problemática relacionada con la longevidad que en otras especies, se trata de un carácter de expresión tardía, por lo que su consideración en los programas de mejora genética alargaría los intervalos generacionales, de ahí que se haya estudiado cómo otros caracteres están relacionados con ella.

En trabajos en los que se estimó la relación fenotípica entre diferentes caracteres morfológicos, especialmente los relacionados con la salud de los aplomos a edades tempranas, y la longevidad se comprobó que estas medidas pudieran ser en cierta medida indicadores de la longevidad de las cerdas (Grindflek y Sehested, 1996; Jørgensen, 1996).

En los estudios genéticos de estimación de correlaciones se puede decir que no se han encontrado caracteres fuertemente correlacionados con la longevidad. López-Serrano y col. (2000) encontraron correlaciones genéticas moderadas entre dos medidas de *stayability* y ganancia diaria (-0.28 y -0.32) y espesor de la grasa dorsal (0.22 y 0.27)

en la raza Large-White. En la raza Landrace las correlaciones con ganancia diaria fueron menores (-0.06 y -0.12), pero, igualmente, las correlaciones fueron moderadas para el espesor de la grasa dorsal (0.24 y 0.11). En esta raza, además, las correlaciones genéticas con los aplomos fueron también moderadas (0.19 y 0.36), cosa que no ocurrió en la raza Large White. Con otros caracteres morfológicos estudiados en este trabajo las correlaciones genéticas fueron muy pequeñas. En el trabajo de Serenius y Stalder (2004), analizando las poblaciones Landrace y Large White finlandesas, los resultados fueron sensiblemente diferentes. Para la raza Landrace los valores más elevados de correlaciones genéticas entre longevidad y otros caracteres se dieron con el número de destetados en el primer parto (0.39), el intervalo entre el primero y el segundo parto (-0.43) y los aplomos (0.32). Para la raza Large White los caracteres que presentaron mayor correlación con la longevidad fueron, al igual que para la raza Landrace, el número de destetados en el primer parto (0.30), el intervalo entre el primer y el segundo parto (-0.40) y los aplomos (0.17); pero, además, presentaron una correlación moderada la edad al primer parto (-0.28) y el espesor de la grasa dorsal a los 100 kg. de peso (0.22). En este trabajo no se detectó correlación negativa entre la longevidad y la ganancia diaria. En el trabajo de Tholen y col. (1996a) se estiman unas correlaciones genéticas para la población sintética entre dos mediadas de *stayability* (STAY13 y STAY14) y el intervalo destete-concepción después del primer parto de -0.45 y -0.22; para esta misma población las correlaciones genéticas estimadas entre estos dos caracteres de *stayability* y el intervalo entre el primer y el segundo parto fueron de -0.54 y -0.25. En otro trabajo de Tholen y col. (1996b) se hace una estimación de las correlaciones genéticas entre las medidas de *stayability* descritas por Tholen y col. (1996a) y otra serie de caracteres productivos (en los mismos dos rebaños). En el rebaño sintético se aprecian correlaciones genéticas negativas entre el número de nacidos vivos en el primer parto y todas las medidas de *stayability* (verdaderas o funcionales) y correlaciones genéticas positivas entre el número de nacidos vivos en los partos segundo y tercero y todas las medidas de *stayability* (verdaderas o funcionales). Para el rebaño formado por hembras Landrace y Large White todas las estimas entre el número de nacidos vivos (en cualquier parto) y todas las medias de *stayability* tienden a ser positivas. De todas las estimas de correlación genética entre prolificidad y *stayability* el valor más alto que se estimó fue 0.45, por lo que se puede decir que estas correlaciones son en general medias-bajas. En general, para ambos rebaños, se obtuvieron estimas de correlaciones genéticas negativas, aunque bajas, entre todas las medidas de *stayability* y

el peso vivo medio de la camada al nacimiento. Igualmente las correlaciones entre las medidas de *stayability* y el crecimiento de la cerda fueron negativas y bajas. En relación con la grasa dorsal en el rebaño sintético las correlaciones genéticas fueron muy próximas a cero, mientras que en el otro rebaño se estimaron unas correlaciones genéticas positivas, aunque bajas.

En los trabajos de Tholen y col. (1996a, b) y López-Serrano y col. (2000) no se tiene en cuenta la naturaleza binaria de las variables objeto de estudio, sino que se analizan con métodos que asumen una distribución Normal para la variable respuesta.

1.3.5.- Especies de laboratorio

En general hay una gran cantidad de estudios realizados en especies de laboratorio que apuntan a que los genes pueden jugar un papel importante en la determinación de la longevidad. De hecho, se han encontrado mutantes que manifiestan una gran longevidad en mosca (*Drosophila melanogaster*), nematodos (*Caenorhabditis elegans*) y ratón (*Mus musculus*). En estos casos por longevidad se entiende, sencillamente, duración de la vida. En general estos mutantes lo son para genes involucrados en procesos endocrinos: péptidos similares a la insulina, factores de crecimiento similares a la insulina, moléculas relacionadas con el metabolismo de la grasa y esteroides (Tatar y col., 2003).

En ratón, además de las evidencias hasta ahora descritas en las especies ganaderas de la determinación genética de la longevidad (existe variabilidad genética entre razas y líneas y es un carácter con heredabilidad diferente de cero), existe una interesante experiencia de selección. En ella se ha obtenido respuesta al seleccionar por longitud de la vida reproductiva (días entre el primero y el último parto, el último parto se definió cuando la hembra no quedó preñada tras 60 días) (Farid y col., 2002). Estos autores estiman unas diferencias entre dos líneas seleccionadas durante 24 generaciones y otra control de 148 y 142 días (5.1 y 5.3 partos) cuando se consideran en el análisis sólo los datos de animales que tras 60 días después del último parto continúan vivos. Si en el análisis se consideran todos los datos (incluyendo datos de animales que mueren antes de llegar al final de su vida reproductiva) las diferencias son de 148 y 105 días (5.3 y 4.4 partos).

1.3.6.- Conejo de carne

1.3.6.1.- Causas de eliminación y estructura de la población

Durante los años 1986 y 1987 se presentaron en los symposia de cunicultura de España una serie de trabajos relacionados con el estudio de las diversas causas de eliminación de las conejas reproductoras en función de diferentes factores, como el seguimiento sanitario, la época del año y el tipo genético. Estos trabajos estaban llevados a cabo en condiciones de granja experimental por lo que sus resultados pudieran no ser extrapolables a condiciones comerciales. El primero de ellos (Torres y col., 1986) consiste en un estudio de las causas de eliminación de los reproductores y de las consecuencias que el seguimiento sanitario de los mismos puede ejercer sobre la sanidad global del núcleo de selección. En cuanto a las causas más comunes de eliminación se citan: los procesos respiratorios, el mal de patas, la muerte del animal y los problemas reproductivos. En relación con el efecto del seguimiento sanitario de la granja sobre las causas de eliminación se aprecia una tendencia decreciente en la eliminación por procesos respiratorios y mal de patas, mientras que las muertes de animales se mantienen o incluso tiende a incrementarse, al igual que la proporción de animales que llegan a las 49 semanas de vida, que es mayor si se hace un seguimiento de las condiciones sanitarias de los animales. En otro estudio de esta serie (Torres y col., 1987a), se observa que hay un efecto de la línea genética a la que pertenece el animal sobre la distribución de las causas de eliminación.

Otro trabajo, en el que se estudian las causas de eliminación de los animales, es el de Canet y col. (2000). Este estudio se llevó a cabo en un núcleo de selección por lo que los resultados pudieran igualmente no ser extrapolables a granjas comerciales. En total se estudiaron 1116 hembras, de ellas un 82 % fueron animales eliminados y un 18 % fueron animales que murieron en producción. Hay que decir que de los eliminados un 59 % eran animales sanos pero que por cuestiones de manejo del núcleo tuvieron que ser sustituidos por los de la siguiente generación. Las causas de eliminación por problemas de salud o muerte más frecuentes, fueron los problemas reproductivos, seguidos de los procesos respiratorios. Dos resultados generales de este trabajo son por un lado que los animales sanos en el momento de su eliminación tuvieron en general mayor producción que los que morían o enfermaban y por otro que los animales sanos tuvieron una media de número de partos mucho mayor, 10.5, que los enfermos o muertos, 3.1 -- 5.9 (variación según las distintas causas de muerte o eliminación).

En un trabajo más reciente y llevado a cabo en granjas comerciales se estudió el estado sanitario de las mismas durante el año 2002 (Rosell, 2003). En cuanto a las proporciones de animales eliminados por distintas causas se comparan tres granjas: una con problemas respiratorios, otra con infección de estafilococos y otra sin ninguna enfermedad predominante. En la tercera, de un total de 1142 animales, un 27 % murieron y un 61.7 % fueron eliminados. De entre los eliminados un 38.1 % lo fueron por baja productividad asociada a problemas reproductivos (p.e.: infertilidad y baja prolificidad), un 19.4 % por otros problemas diferentes a los contemplados como categorías principales y un 16.9 % por su condición corporal.

En el 2003 las principales causas de eliminación en condiciones comerciales fueron, problemas reproductivos, mamitis y mal de patas (J. M. Rosell comunicación personal).

Un dato de interés que hay que resaltar en el trabajo de Rosell (2003) es que, a veces, en las granjas objeto de estudio (en total 110) el 50 % de las hembras muertas o eliminadas lo eran en el curso de sus tres primeros partos. Un resultado muy similar a éste fue obtenido previamente para animales de un núcleo de selección por Torres y col. (1987b).

Desde 1991 hasta el 2000, la tasa de reposición en la cunicultura de carne española se ha mantenido prácticamente constante alrededor del 120 % (Ramón y Rafel, 2002). Esto supone, si se mantiene constante el tamaño de la población y su estructura de edades, una esperanza de vida (desde el inicio de la recría) de 304 días. En los resultados de gestión técnico-económica del año 2000 para España el intervalo entre partos se cifra en 43.3 días (Ramón y Rafel, 2002), por lo que esa esperanza de vida equivaldría a 5.3 partos, si se contempla un periodo de recría de 2.5 meses.

Una consecuencia directa de esta elevada tasa de reposición es que la producción se sustenta en animales jóvenes. Este hecho es mostrado explícitamente por Rosell y col. (1996) en un estudio de 94 granjas, llevado a cabo durante 1992 y 1993. En este trabajo de 4719 animales controlados, 682 (14 %) están en su primer ciclo, 641 (13.5 %) en el segundo, 558 (11.8 %) en el tercero y en ciclos superiores los porcentajes siguen decreciendo. El que la producción se sustente sobre animales jóvenes puede tener una consecuencia en los rendimientos productivos pues se ha constatado una menor productividad de estos animales (Gómez, 1994). Este autor da las medias para número de nacidos vivos, destetados y sacrificados en los tres primeros partos en dos

líneas seleccionadas por caracteres reproductivos, observándose un comportamiento inferior del primero con respecto al segundo y tercero.

1.3.6.2.- Importancia económica de la longevidad o de la tasa de reposición

En general, los mismos argumentos de carácter cualitativo, que se han dado en otras especies, se pueden dar en conejo para justificar el estudio y la consideración de la longevidad en los programas de mejora genética. Al aumentar la longevidad y por tanto reducir la tasa de reposición se reducirán los costes directos derivados de la reposición; se podrá cambiar la estructura de la población de manera que la producción se sustente en animales de las clases de edad más productivas; si la reposición se hace con animales del exterior el flujo de animales que deben de entrar en la granja se reducirá por lo que se evitarán problemas de falta de adaptación a las condiciones específicas de cada granja y, por último, si se mejoran las aptitudes funcionales de las conejas (son más resistentes a los problemas patológicos) la proporción de animales que se pueden eliminar por cuestiones meramente productivas puede ser mayor, por lo que aumentaría la eficacia en la producción. A pesar de los aspectos positivos que implicarían un mejora de la longevidad de las hembras y por tanto una menor tasa de reposición, la consideración del carácter como un objetivo en programas de selección intralínea no parece ser muy adecuada pues ello implicaría, dadas las peculiaridades del carácter, largos intervalos generacionales, lo que a su vez se traduciría en repuestas muy bajas por unidad de tiempo.

Solamente existe en la bibliografía un trabajo en el que se cuantifican los pesos económicos de los caracteres productivos en cunicultura para la situación de España (Armero y Blasco 1992). En este estudio se estima que un descenso de 1% en la tasa de reposición anual supone un incremento en los beneficios de 0.34 Euros por coneja y año. En este trabajo los tres caracteres con mayor peso económico fueron el número de partos al año, el tamaño de camada y la ganancia media diaria durante el periodo de cebo. La tasa de reposición se situó en octavo lugar después del consumo de pienso en lactación. El peso económico de cada uno de los caracteres estudiados en este trabajo tiene sus propias unidades, (por ejemplo euros por aumento de un parto al año o euros por descenso de un 1% en la tasa de reposición) lo que hace que la comparación entre caracteres no sea inmediata. Una manera de expresar estos resultados de manera que todos tengan las mismas unidades y puedan ser comparables es hacerlo por ejemplo en euros por aumento de una desviación típica genética del carácter. Si se hiciera de esta

forma muy probablemente se reduciría el peso económico del número de partos al año, dada su baja variabilidad genética (Baselga y Blasco, 1989). Otra observación a los resultados de este trabajo es el hecho de que en la obtención de los costes de reposición de una hembra la situación actual puede ser diferente a la que se plasma en este trabajo (situación de 1992). Actualmente el cunicultor no percibe nada por los animales de desvieje; es más, si se mueren tendrá que pagar por el procesamiento de los cadáveres. Otro punto que creemos es muy importante y no es tenido en cuenta en este trabajo es que la mayor parte de los animales que se renuevan lo son porque caen enfermos (incluyendo como enfermos las conejas eliminadas por infertilidad) o mueren. Estos animales van a ser individuos que desde que caen enfermos hasta que finalmente son repuestos pueden haberse mantenido en la granja con un nivel de producción más bajo que el de otro animal sano. Puede que con su mantenimiento en la granja no se incurra en gastos extras (si no se trata médicamente) pero si en un descenso en los beneficios.

1.3.6.3.- Variabilidad, parámetros genéticos y relación con la producción

Piles y col. (2005) han apreciado un diferente comportamiento en cuanto a longevidad, definida como el número de días desde el primer test de gestación positivo hasta la muerte o eliminación, para los diferentes tipos genéticos involucrados en un cruce dialélico completo entre tres líneas maternas. Las diferencias entre los tipos puros se deben a diferencias entre los efectos genéticos directos, mientras que las diferencias entre tipos puros y cruzados se atribuyen además de a las diferencias entre los efectos genéticos mencionados a efectos de heterosis, que para algunos de los cruces de este experimento fueron significativos. En este trabajo no se detectaron efectos genéticos maternos como significativos y su magnitud en general fue baja. La diferencia entre los efectos maternos suponía entre un 57% y un 16% de las diferencias entre los efectos genéticos directos. Otro trabajo en el que se comparan las producciones acumuladas de diferentes razas de Estados Unidos (Neozelandés y California) y su cruce es el de Lukefahr y Hamilton (2000). En este estudio tanto para caracteres de productividad acumulada como para la propia longevidad la raza Neozelandesa se mostró superior a la California y los animales cruzados se comportaron de manera similar a los Neozelandeses. Otro trabajo más antiguo en el que se hace un estudio comparativo de las tasas de mortalidad y eliminación, y las causas de muerte y eliminación, de hembras pertenecientes a los cuatro tipos genéticos procedentes de un cruce dialélico completo entre animales de origen Neozelandés y California es el de

Coudert y Brun (1988). Los resultados se expresan en términos del modelo de Dickerson y no se observan diferencias entre los efectos genéticos directos ni entre los efectos genéticos maternos ni para la tasa de mortalidad ni para la tasa de eliminación. En cuanto a la heterosis se estiman valores favorables y significativos para la tasa de mortalidad y también significativos aunque desfavorables para la tasa de eliminación. Por lo tanto una primera conclusión asociada a los resultados de estos tres trabajos es que en conejo existe variabilidad genética para la longevidad entre razas y líneas.

Solamente se han encontrado dos trabajos en la bibliografía que estudian la heredabilidad de la longevidad. El primero es el de Youssef y col. (2000), en el que la metodología empleada para el análisis es de modelo lineal mixto, pero no se da información sobre cómo se tratan los datos censurados (a las hembras no se les permitía producir más de 10 camadas). Probablemente estos datos se tratasen como completos. Además, el modelo empleado para estudiar la componente genética del carácter es un modelo macho. En general, en las poblaciones de conejos las premisas de aplicación del modelo macho, de tal manera que tenga un modelo animal equivalente acorde con los postulados del modelo genético infinitesimal, suelen estar lejos de ser cumplidas. En primer lugar hay individuos que producen dato y que además aparecen en la genealogía como genitores; las hembras no sólo están emparentadas a través de su padre y tampoco se cumple la condición de que las hijas de un abuelo materno que están representadas en los datos son una muestra aleatoria de todas las hijas de este abuelo. Tampoco se cumple que las hijas se apareen con un solo macho y tengan un único descendiente, ni que el número de hembras con el que se aparea cada macho sea lo suficientemente grande como para que formen una muestra aleatoria y representativa de la población de hembras. Independientemente de la validez del método y del modelo en este trabajo, la estima de la heredabilidad para la longevidad, definida como el número de meses que dura la vida productiva, fue de 0.13. El otro trabajo en el que se hacen estimaciones de la heredabilidad es el de Garreau y col. (2001), en este estudio el método de análisis para estudiar la longevidad, medida como el número de inseminaciones a que se somete una coneja hasta que es eliminada o muere, es el análisis de supervivencia. En este trabajo igualmente se aborda la estimación de la componente genética del carácter con modelos macho, y como se ha comentado anteriormente las premisas de aplicación de estos modelos en cunicultura suelen estar lejos de cumplirse. Los valores estimados en este trabajo para la heredabilidad efectiva varían entre 0.24 cuando no se tiene en

consideración las relaciones de parentesco entre los machos y el carácter se estudia como continuo, y 0.05 cuando se considera las relaciones de parentesco entre los machos y el carácter se estudia usando una versión discreta de la metodología de análisis de supervivencia.

En cuanto a su relación con otros caracteres de producción (prolificidad) son pocos los trabajos que se han encontrado en la bibliografía y en ninguno de ellos se aborda directamente la estimación de las correlaciones entre ellos, sino que se trata de estudios en los que se estima la respuesta correlacionada en longevidad tras un proceso de selección por prolificidad. El más antiguo es el de Rinaldo y Bolet (1988), en este trabajo se comparan dos tipos genéticos: uno perteneciente a una línea seleccionada durante 7 generaciones por prolificidad en los tres primeros partos y otro formado por la población control de la línea seleccionada. No se encontraron diferencias significativas ni para el número de partos ni para la longitud de la vida productiva (días desde la primera hasta la última camada) entre estos dos tipos de animales, aunque la magnitud de las diferencias entre los grupos era relevante (1.1 partos a favor de la línea control y 73 días igualmente a favor de la línea control). En este estudio hay datos censurados (28%) pero no se indica de manera explícita cómo se tratan, aunque por la metodología empleada para el análisis (modelo lineal) es de suponer que los datos de estos animales se trataron como datos completos. El otro trabajo que aborda la estimación de la respuesta correlacionada en longevidad después de seleccionar por prolificidad es el de Theilgaard y col. (2005). Se compara la longevidad, medida como el número de días desde el primer test de gestación positivo hasta la muerte o eliminación por problemas de salud de las conejas, en dos grupos de animales cruzados distintos. La diferencia entre los dos tipos de animales era que unos procedían del cruce actual de dos líneas seleccionadas por tamaño de camada y otros de un cruce de animales de 12 generaciones atrás, que hasta el momento de realizar la comparación se mantuvieron como embriones criopreservados. En este trabajo la metodología empleada es el análisis de supervivencia y se obtiene que la eliminación o muerte de un animal “antiguo” es 1.6 veces más probable que la de uno “actual”. Este efecto, a pesar de ser relevante, dado que el número de animales en el experimento era pequeño, no alcanzó significación estadística.

De acuerdo con la clasificación que hace Essl (1998) de los caracteres en función de su importancia en la determinación de la *fitness*, los caracteres de longevidad y de tamaño de camada se agrupan en la categoría que mayor influencia tienen en la

fitness. Por lo tanto en conejo el efecto de la selección por tamaño de camada pudiera no tener consecuencias negativas sobre la longevidad, pues ambos son componentes importantes y que actúan de la misma manera en la determinación de la *fitness*. Sin embargo en vacuno lechero, la selección por producción lechera (un carácter que tiene un óptimo intermedio para la *fitness*) ha deteriorado la longevidad funcional. Concordantes con el anterior razonamiento pueden ser los resultados de los dos trabajos anteriormente comentados (Theilgaard y col., 2005; Rinaldo y Bolet, 1988) en los que no se detectan diferencias significativas para longevidad entre los grupos seleccionados por prolificidad y sus respectivos grupos controles, por lo que es de esperar que la correlación entre ambos caracteres sea baja.

1.4.- Conclusiones

En general para todas las especies la longevidad se puede considerar como un carácter de heredabilidad baja, a pesar de que la interpretación de este parámetro no es equivalente entre las diferentes definiciones del carácter ni entre los distintos métodos de análisis. En la tabla 1.4.1 se resumen por especies las distintas estimas que se han obtenido para este parámetro. Las correlaciones genéticas entre diferentes medidas de longevidad y otros caracteres de interés se pueden considerar como medias-bajas. A pesar de la amplia bibliografía que trata el carácter en la mayor parte de las especies ganaderas, hasta ahora, únicamente en vacuno lechero se ha contemplado el carácter como un objetivo directo de selección, para el que se hacen rutinariamente evaluaciones genéticas.

Tabla 1.4.1.- Resumen de la estimas de heredabilidades en las distintas especies ganaderas

	Referencia	Estimas ^a	Método	Observaciones
Vacuno lechero	Chirinos (2003) ^b	0.078 -- 0.298	A.S. ^c	heredabilidad efectiva
	Ducrocq (1987) ^b	0.008 -- 0.39	Diversos	
Vacuno de carne	Silva y col. (2003)	0.11 -- 0.17	Stay. ^d	Considerado de naturaleza binaria
	Doyle y col. (2000)	0.15	Stay.	Considerado de naturaleza binaria
	Snelling y col. (1995)	0.025 -- 0.23	Stay.	Considerado de naturaleza binaria
	Díaz y col. (2002)	0.031 -- 0.041	A.S.	heredabilidad en escala original
	Hyde (2000)	0.053	A.S.	heredabilidad en escala logarítmica
Ovino lechero	El-Saied y col. (2005)	0.02 -- 0.06	REML	No incluye datos censurados
Cerdos	Yazdi y col. (2000a)	0.11 -- 0.27	A.S.	heredabilidad en escala original
	Yazdi y col. (2000b)	0.21 -- 0.31	A.S.	heredabilidad en escala original
	Serenius y Stalder (2004)	0.17 -- 0.19	A.S.	heredabilidad en escala logarítmica
	Guo y col. (2001)	0.25	MLM ^e con censura	heredabilidad en escala original
	López-Serrano y col. (2000)	0.08 -- 0.11	Stay. (REML)	No considerado de naturaleza binaria
	Tholen y col. (1996a)	0.02 -- 0.09	Stay. (REML)	No considerado de naturaleza binaria
	Conejos	Youssef y col. (2000)	0.13	REML
Garreau et al. (2001)		0.05 -- 0.24	A.S.	heredabilidad efectiva

^a Dentro de un trabajo los rangos indican diferentes poblaciones estudiadas o estimas obtenidas usando diferentes métodos o definiciones del carácter, ^b Revisión, ^c Análisis de Supervivencia, ^d *Stayability*, ^e Modelo Lineal Mixto

Con respecto a los aspectos metodológicos del análisis del carácter, el análisis de supervivencia ofrece un marco idóneo para su estudio, pues permite considerar con absoluto rigor la información contenida en los datos censurados, modelar covariables

con efecto dependiente del tiempo y considerar la naturaleza no Normal de la medida fenotípica. Además desde un punto de vista teórico ya se han resuelto aspectos metodológicos que inicialmente eran puntos negativos en contra de esta metodología, modelos multicarácter y modelado del efecto genético como dependiente del tiempo. A pesar de las virtudes estadísticas de las técnicas de análisis de supervivencia para el modelado de datos de longevidad, estos métodos presenta el importante inconveniente de los elevados tiempos de cómputo requeridos para su implementación. Actualmente parece existir consenso para abordar el análisis de datos de longevidad usando estas técnicas, la única alternativa a ellos que se puede considerar como aceptable podría ser el uso de modelos de regresión aleatoria para estudiar los “códigos de supervivencia”, aunque como ya se ha comentado, de esta manera se estarían empleando métodos lineales para modelar variables de respuesta binaria, lo que podría no ser óptimo desde un punto de vista estadístico.

Con respecto a nuestro trabajo hemos elegido el marco metodológico que ofrece las técnicas de análisis de supervivencia para por un lado, abordar la estimación de parámetros genéticos de la longevidad en conejo de producción cárnica, pues la literatura a este respecto es prácticamente nula, y por otro, para evaluar la longevidad de una línea que se ha constituido aplicando grandes presiones de selección en poblaciones comerciales para longevidad y con niveles independientes alrededor de la media para prolificidad. Para la constitución de la nueva línea se ha optado por utilizar métodos de selección por “hiperlongevidad”, tras considerar por un lado los resultados obtenidos en programas y experimentos de selección por “hiperprolificidad” realizados en cerdos y conejos, y por otro las dificultades intrínsecas del carácter longevidad (expresión tardía, baja heredabilidad y distribución fenotípica no Normal) que harían esperar respuestas muy bajas en programas de selección intra-línea.

1.5.-Bibliografía

- Andersen A.H., Korsgaard I.R., Jensen J., 2000. Identifiability of parameters in – and equivalence of animal and sire models for gaussian and threshold characters, traits following a poisson mixed model, and survival traits. Research Report no. 417, Department of theoretical statistics. University of Aarhus. (Denmark).
- Armero Q., Blasco A., 1992. Economic Weights for rabbit selection indices. *J. Appl. Rabbit Res.* 15:637-642.
- Arthur P.F., Makarechian M., Berg R.T., Weingardt R., 1993. Longevity and Lifetime productivity of Cows in a purebred Hereford and two multibreed synthetic groups under range conditions. *J. Anim. Sci.* 71:1142-1147.
- Baselga M., Blasco A., 1989. Mejora genética del conejo de producción de carne. Ediciones Mundi Prensa, Madrid, (Spain).
- Beamonte E., Bermúdez J.D., 2003. A Bayesian semiparametric analysis for additive hazard models with censored observations. *Sociedad de Estadística e Investigación Operativa. Test* 12:101-117.
- 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. In: Proc. of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier (France), Vol. 32 pp. 509 - 512.
- Brandt H., von Brevern N., Glodek P., 1999. Factors affecting survival rate of crossbred sows in weaner production. *Livestock Production Science* 57:127-135.
- Breslow N.E., 1974. Covariance analysis of censored survival data. *Biometrics* 30:89-99.

- Boettcher P.J., Jairath L.K., Dekkers J.C.M., 1999. Comparison of methods for 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., 1997. Genetic parameters for simple predictor of lifespan of Holstein-Friesian dairy cattle and its relationship to production. *Animal Science* 65:31-37.
- Canet M., Santacreu M.A., Torres C., 2000. Effect of disease processes on reproductive performance of rabbit does. In: Proc. of the 7th World Rabbit Congress, Valencia (Spain). Vol B, pp. 407-412.
- Casas E., Freking B.A., Leymaster K.A., 2004. Evaluation of Dorset, Finnsheep, Romanov, Texel, and Montadale breeds of sheep: II. Reproduction of F1 ewes in fall mating seasons. *J. Anim. Sci.* 82:1280-1289.
- Collett D., 2003. *Modelling Survival data in medical research*, 2nd Edition. Chapman & Hall/CRC Press LLC (USA).
- Coudert P., Brun J.M., 1988. Production et morbidité des lapines reproductrices: étude comparative de quatre génotypes. *Genet. Sel. Evol.* 21:49-65.
- Cox D.R., 1972. Regression models and life tables (with discussion). *J. R. Stat. Soc. B* 34:187-220.
- Chirinos Z.R., 2003. Desarrollo de un sistema de valoración genética para la longevidad en ganado frisón español. Tesis doctoral. Universidad Politécnica de Madrid (Spain).
- Damgaard L.H., 2004. Quantitative genetic analysis of survival, linear Gaussian and ordered categorical traits. Ph.D. Thesis. The Royal Veterinary and Agricultural University, Frederiksberg C (Denmark).

- Damgaard L.H., Korsgaard I.R., Simonsen J., Dalsgaard O., Andersen A.H., 2003. Genetic versus non-genetic Weibull log-normal sire frailty models: A simulation study. In: Proc. of the 2003 interbull technical workshop, Belstville, (USA). Bulletin No 30, pp. 35-42.
- Dekkers J.C.M., 1993. Theoretical basis for genetic parameters of herd life and effects on response to selection. *J. Dairy Sci.* 76:1433-1443.
- Dekkers J.C.M., Jairath L.K., 1994. Requirements and uses of genetic evaluations for conformation and herd life. In: Proc. of the 5th World Congress on Genetics Applied to Livestock Production, Univ. of Guelph (Canada), Vol. 17 pp. 61-68.
- Diaz C., Chirinos Z., Moreno A., Carabaño M.J., 2002. Preliminary analysis of functional longevity in the Avileña Negra Ibérica beef cattle breed. In: Proc. of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier (France), Vol. 29 pp. 697-700.
- Doyle S.P., Golden B.L., Green R.D., Brinks J.S., 2000. Additive genetic parameter estimates for heifer pregnancy and subsequent reproduction in Angus females. *J. Anim. Sci.* 78:2091-2098.
- Ducrocq V., 1987. An analysis of length of productive life in dairy cattle. Ph.D. Thesis. Cornell University, Ithaca, New York (USA).
- Ducrocq V., 1999a. Two years of experience with the French genetic evaluation of dairy bulls on production-adjusted longevity of their daughters. In: Proc. of the international workshop on EU concerted action, genetic improvement of functional traits in cattle (GIFT); Longevity, Jouy-en-Josas, (France). Bulletin No 21, pp. 60-67.
- Ducrocq V., 1999b. Topics that may deserve further attention in survival analysis applied to dairy cattle breeding – some suggestions. In: Proc. of the international workshop on EU concerted action, genetic improvement of functional traits in cattle (GIFT); Longevity, Jouy-en-Josas, (France). Bulletin No 21, pp. 181-189.

- Ducrocq V. 2001a. A two step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of productive life. In: Proc. of the Interbull meeting, Budapest, (Hungary). Bulletin No 27, pp. 147-152.
- Ducrocq V., 2001b. Survival analysis applied to animal breeding and epidemiology. Course notes.
- Ducrocq V., 2005. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. *Animal Science* 80:249-255.
- Ducrocq V., Casella G., 1996. A bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 28:505-529.
- Ducrocq, V., Delaunay, I., Boichard, D., Mattalia S., 2003. A general approach for international genetic evaluations robust to inconsistencies of genetic trends in national evaluations. In: Proc. of the 2003 Interbull meeting. Beltsville, (USA). Bulletin No 30, pp. 101-111.
- Ducrocq V., Quaas R.L., Pollak E.J., Casella G., 1988a. Length of productive life of dairy cattle. 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 cattle. 2. Variance components estimation and sire evaluation. *J. Dairy Sci.* 71:3071-3079.
- Ducrocq V, Sölkner J, 1994. "The survival Kit". A fortran package for the analysis of survival data. In: Proc. of the 5th World Congress on Genetics Applied to Livestock Production, University of Guelph, (Canada), Vol. 22 pp. 51.
- El-Saied U.M., De la Fuente L.F., Carriedo J.A., San Primitivo F., 2005. Genetic and phenotypic parameter estimates of total and partial lifetime traits for dairy ewes. *J. Dairy Sci.* 88:3265-3272.

- Essl A., 1998. Longevity in dairy cattle breeding: a review. *Livestock Production Science* 57:79-89.
- Everett R.W., Keown J.F., Clapp E.E., 1976a. Relationships 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.
- Famula T.R., 1981. Exponential stayability model with censoring and covariates. *J. Dairy Sci.* 64:538-545.
- Farid A., Crober D.C., Van der Steen H., Patterson D.L., Sabour M.P., 2002. Reproductive performance of mice selected for reproductive longevity. In: *Proc. of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier (France), Vol. 30 pp. 681-684.*
- Garreau H., Larzul C., Ducrocq V., 2001 Analyse de la longévité de la souche de lapins INRA 1077. In: *Proc. of the 9^{èmes} Journ. Rech. Cunicole, Paris (France).* pp. 217 – 220.
- Gianola D., 1980. A method of sire evaluation for dichotomies. *J. Anim. Sci.* 51:1266-1271.
- Gómez E.A., 1994. La selección del tamaño de camada en conejo de carne : Influencia de los efectos maternos y de la heterogeneidad genética entre partos. Tesis doctoral. Universidad Politécnica de Valencia (Spain).
- Grindflek E., Sehested E., 1996. Conformation and longevity in Norwegian pigs. In: *Proc. of NJF (Nordiska Jordbruksforskarens Förening) Seminar no. 265. Research Centre Foulum, (Denmark).* pp. 77-84.

- Guo S-F., Gianola D., Rekaya R., Short T., 2001. Bayesian analysis of lifetime performance and prolificacy in Landacre sows using a linear mixed model with censoring. *Livestock Production Science* 72:243-252.
- Hall A.D., Lo S., Rance K. A., 2002. Comparative study of the lifetime productivity and performance characteristic of Meishan and Duroc cross-bred pigs. *Acta Agric. Scand., Sect. A, Animal Sci.* 52:183-188.
- Höschele I., Gianola D., Foulley J.L., 1987. Estimation of variance components with quasi-continuous data using Bayesian methods. *Journal of Animal Breeding and Genetics* 104:334-349.
- Hyde L.R., 2000. An application of survival analysis to the genetic evaluation of reproductive life of beef females. Ph.D. Thesis. Colorado State University, Fort Collins, Colorado (USA).
- 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.
- Jensen J., Korsgaard I.R., Neerhof H.J., Vollema A., Madsen P., Ducrocq V., 1999. Genetic variation in functional longevity and its relation to mastitis resistance in Danish Holstein. In: Proc. of the international workshop on EU concerted action, genetic improvement of functional traits in cattle (GIFT); Longevity, Jouy-en-Josas, (France). Bulletin No 21, pp. 161-165.
- Jørgensen B., Sørensen M. T., 1998. Different rearing intensities of gilts: II. Effects on subsequent leg weakness and longevity. *Livestock Production Science* 54:167-171.
- Jørgensen B., 1996. The influence of leg weakness in gilts, on their longevity as sows, assessed by survival analysis. In: Proc. of NJF (Nordiska Jordbruksforskarens Förening) Seminar no. 265. Research Centre Foulum, (Denmark). pp. 95-100.

- Jungst S.B., Kuhlers D.L., Little J.A., 1988. Longevity and maternal productivity of F1 crossbred Landrace Sows managed in two different gestation systems. *Livestock Production Science* 19:499-510.
- Kachman S.D., 1999. Applications in survival analysis. *J. Anim. Sci.* 77(Supplement 2): 147-153.
- Kalbfleisch J. D., Prentice R. L., 1980. *The statistical analysis of failure time data.* John Wiley and Sons, New York (USA).
- Kimura F., Fujii T., 2003. Globalizing Activities and the Rate of Survival: Panel Data Analysis on Japanese Firms. NBER Working Paper No. W10067.
- 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.L., 1997. *Survival analysis techniques for censored and truncated data.* Springer-Verlag, New York (USA).
- Korsgaard I.R., Andersen A.H., Jensen J., 1999. Discussion of heritability of survival traits. In: Proc. of the international workshop on EU concerted action, genetic improvement of functional traits in cattle (GIFT); Longevity, Jouy-en-Josas, (France). Bulletin No 21, pp. 31-35.
- 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.
- Korsgaard I.R., Lund M.S., Sorensen D., Gianola D., Madsen P., Jensen J., 2003. Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genet. Sel. Evol.* 35:159-183.

- 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.
- Kühn Ch., Bennewitz J., Reinsch N., Xu N., Thomsen H., Looft C., Brockmann G.A., Schwerin M., Weimann C., Hiendleder S., Erhardt G., Medjugorac I., Förster M., Brenig B., Reinhardt F., Reents R., Russ I., Averdunk G., Blümel J., Kalm E., 2003. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J. Dairy Sci.* 86:360-368.
- Larroque H., Ducrocq V., 2001. Relationships between type and longevity in the Holstein breed. *Genet. Sel. Evol.* 33:39-59.
- López-Serrano M., Reinsch N., Looft H., Kalm E., 2000. Genetic correlations of growth, backfat thickness and exterior with stayability in Large White and Landrace sows. *Livestock Production Science* 64:121-131.
- Lucia T., Dial G.D., Marsh W.E., 2000. Lifetime reproductive performance in female pigs having distinct reasons for removal. *Livestock Production Science* 63:213-222.
- Lukefahr S.D., Hamilton H.H., 2000. Longevity and cumulative litter productivity in straightbred and crossbred Californian and New Zealand White does. In: *Proc. of the 7th World Rabbit Congress, Valencia (Spain)*. Vol A, pp. 463-468.
- Madgwick P.A., Goddard M.E., 1989. Genetic and phenotypic parameters of longevity in Australian dairy cattle. *J. Dairy Sci.* 72:2624-2632.
- Mrode R.A., 1996. *Linear models for the prediction of animal breeding values*. CAB International, (UK).
- Núñez-Dominguez R., Cundiff L.V., Dickerson G.E., Gregory K.E., Koch R.M., 1991. Heterosis for survival and dentition in Hereford, Angus, Shorthorn and crossbred cows. *J. Anim. Sci.* 69:1885-1898.

- O'Dowd S., Hoste S., Mercer J.T., Fowler V.R., Edwards S.A., 1997. Nutritional modification of body composition and the consequences for reproductive performance and longevity in genetically lean sows. *Livestock Production Science* 52:155-165.
- Péres-Razo M., Sánchez F., Torres-Hernández G., Becerril-Pérez C., Gallegos- Sánchez J., González-Cosío F., Meza-Herrera C., 2004. Risk factors associated with dairy goats stayability. *Livestock Production Science*, 89:139-146.
- Piles M., Sánchez J.P., Orengo J., Rafel O., Ramon, J., Baselga, M., 2005. Crossbreeding parameters estimation for functional longevity in rabbits using survival analysis methodology. *J. Anim. Sci.* (*in press*).
- Ramón J., Rafel O., 2002. 1991-2000. Diez años de gestión global en España. In: Proc. of Expoaviga 2002, X jornada cunícola, Barcelona (Spain), pp.113–117.
- Rendel J.M., Robertson A., 1950. Some aspects of longevity in dairy cows. *Empire J. Exp. Agric.* 18 (69): 49-56.
- Rinaldo D., Bolet G., 1988. Effect of selection for litter size at weaning on reproductive life of female rabbits. In: Proc. of the 4th World Rabbit Congress, Budapest (Hungary). Vol. 1, pp. 269-275.
- Rodriguez-Zas S.L., Southey B. R., Knox R.V., Connor J.F., Lowe J.F., Roskamp B.J. 2003. Bioeconomic evaluation of sow longevity and profitability. *J. Anim. Sci.*, 81:2915-2922.
- Rosell J.M., 2003. Health status of commercial rabbits in the Iberian peninsula. A practitioner's study. *World Rabbit Sci.* 11:157-169.
- Rosell J.M., Badiola J.I., De la Fuente L.F., 1996. Rhinitis of the domestic rabbit. An epidemiological survey during the period 1992-1993. Influence of the sex, the age and the phase of lactation. In: Proc. of the 6th World Rabbit Congress, Toulouse, (France). Vol 3, pp. 113-117.

- Roxstöm A., Ducrocq V., Strandberg E., 2003. Survival analysis of longevity in dairy cattle on a lactation basis. *Genet. Sel. Evol.* 35:305-318.
- Schaeffer L.R., 2004. Application of random regression models in animal breeding. *Livestock Production Science* 86:35-45.
- Sehested E., 1996. Economy of sow longevity. In: Proc. of NJF (Nordiska Jordbruksforskarens Förening) Seminar no. 265. Research Centre Foulum, (Denmark). pp. 101-108.
- Serenius T., Stalder K.J., 2004. Genetics of length of productive life and lifetime prolificacy in the Finnish Landrace and Large White pig populations. *J. Anim. Sci.* 82:3111-3117.
- Silva J.A.II.V., Eler J.P., Ferraz J.B.S., Golden B.L., Oliveira H.N., 2003. Heritability estimate for stayability in nelore cows. *Livestock Production Science* 79:97-101.
- Silvestre M. A., Salvador I., Sánchez J. P., Gómez E.A., 2004a. Effect of changing female stimulus on intensive semen collection in young Murciano-Granadina male goats. *J. Anim. Sci.* 82:1641-1645.
- Silvestre M. A., Sánchez J. P., Gómez E.A., 2004b. Vitrification of goat, sheep, and cattle skin samples from whole ear extirpated after death and maintained at different storage times and temperatures. *Cryobiology* 49:221-229.
- Srivongs D., Liawnoraset W., Pongskul C., Reungjui S., 2004. Graft survival analysis in kidney transplantation: A 12-year experience in a Thai medical center. *Transplantation Proceedings* 36(7): 2034-2037.
- Smith S.P., Quaas R.L., 1984. Productive lifespan of bull progeny groups: Failure time analysis. *J. Dairy Sci.* 67:2999-3007.

- Smith S.P., Allaire F.R., 1986. Analysis of failure times measured on dairy cows: Theoretical considerations in animal breeding. *J. Dairy Sci.* 69:217-227.
- Snelling W.M., Golden B.L., Bourdon R.M., 1995. Within-Herd genetic analyses of stayability of beef females. *J. Anim. Sci.* 73:993-1001.
- Sölkner J., Petschina R., 1999. Relationship between type traits and longevity in Austrian Simmental cattle. In: Proc. of the international workshop on EU concerted action, genetic improvement of functional traits in cattle (GIFT); Longevity, Jouy-en-Josas, (France). Bulletin No 21, pp. 91-95.
- Strandberg E., Roth A., 1999. Genetic parameters of functional and fertility-determined length of productive life in Swedish dairy cattle. In: Proc. of the international workshop on EU concerted action, genetic improvement of functional traits in cattle (GIFT); Longevity, Jouy-en-Josas, (France). Bulletin No 21, pp. 152-160.
- Strandberg E., Sölkner J., 1996. Breeding for longevity and survival in dairy cattle. In: Proc. of the international workshop on genetic improvement of functional traits in cattle, Gembloux (Belgium). Bulletin N° 12, pp. 111-119.
- Tantasuparuk W., Lundeheim N., Dalin A.-M., Kunavongkrit A., Einarsson S., 2001. Weaning-to-service interval in primiparous sows and its relationship with longevity and piglet production. *Livestock Production Science* 69:155-162.
- Tarres J., Piedrafita J., Ducrocq V., 2005. Validation of an approximate approach to compute genetic correlations between longevity and linear traits. *Genet. Sel. Evol.* (*in press*).
- Tatar M., Bartke A., Antebi A., 2003. The endocrine regulation of aging by insulin-like signals. *Science* 299:1346-1351.
- Theilgaard P., Sánchez J.P., Pascual J.J., Friggens N.C., Baselga M., 2005. Effect of selection for prolificacy, and body fatness on survival of rabbit does assessed

using a cryopreserved control population. Submitted to Livestock Production Science.

Tholen E., Bunter K.L., Hermes S., Graser H-U., 1996a. The genetic foundation of fitness and reproduction traits in Australian pig populations 1. Genetic parameters for weaning to conception interval, farrowing interval, and stayability. *Aust. J. Agric. Res.* 47:1261-1274.

Tholen E., Bunter K.L., Hermes S., Graser H-U., 1996b. The genetic foundation of fitness and reproduction traits in Australian pig populations 2. Relationships between weaning to conception interval, farrowing interval, stayability, and other common reproduction and production traits. *Aust. J. Agric. Res.* 47:1275-1290.

Torres C., Garces M., Fabado F., Pla M., 1987a. Causas de eliminación de reproductores en función de la línea y época. In: *Proc. of the XII Symposium de cunicultura, Guadalajara, (Spain)*, pp. 237-244.

Torres C., Pla M., García F., 1986. Nivel de respuesta en el tiempo a un control de seguimiento sanitario en conejos. In: *Proc. of the XI Symposium de cunicultura, Teruel, (Spain)*, pp. 145-152.

Torres C., Pla M., Garces M., Fabado F., 1987b. Efectos de la intensidad de producción sobre la esperanza de vida media productiva de las conejas. In: *Proc. of the XII Symposium de cunicultura, Guadalajara, (Spain)*, pp. 289-299.

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

VanRaden P.M., Powell R.L., 2002. Properties of international longevity evaluation and correlations with other traits. In: *Proc. of the 2002 Interbull meeting, Interlaken, (Switzerland)* 29:61-65.

Van Vleck L.D., 1980. Stayability evaluation as a categorical trait and by considering other traits. *J. Dairy Sci.* 63:1172-1180.

- Veerkamp R.F., Brotherstone S., Engel B., Meuwissen T.H.E., 2001. Analysis of censored survival data using random regression models. *Animal Science* 72:1-10.
- Vega-Murillo V.E., Kachman S.D., Cundiff L.V., Van Vleck L.D. 1999. Analysis of herd life in a four-breed diallel in beef cattle. *J. Anim. Sci.* 77 (Supplement 1):144.
- Visscher P.M., Goddard M.E., 1995. Genetic parameters for milk yield, survival, workability, type traits for Australian dairy cattle. *J. Dairy Sci.* 78:205-220.
- Visscher P., Thompson R., Yazdi H., Hill W.G., Brotherstone S., 1999. Genetic analysis of longevity data in the UK: present practice and considerations for future. In: Proc. of the international workshop on EU concerted action, genetic improvement of functional traits in cattle (GIFT); Longevity, Jouy-en-Josas, (France). Bulletin No 21, pp. 16-22.
- De Vries A.G., 1989. A model to estimate economic values of traits in pig breeding. *Livestock Production Science* 21:49-66.
- Wolfová M., Wolf J., Zahrádková R., Pribyl J., Dano J., Krupa E., Kica J., 2005. Breeding objectives for beef cattle used in different production systems 2. Model application to production systems with the Charolais breed. *Livestock Production Science* 95: 217-230.
- Yazdi M. H., Lundeheim, N., Rydhmer L., Ringmar-Cederberg E., Johansson K., 2000a. Survival of Swedish Landrace and Yorkshire sows in relation to osteochondrosis: a genetic study. *Animal Science* 71:1-9.
- Yazdi M. H., Rydhmer L., Ringmar-Cederberg E., Lundeheim, N., Johansson K., 2000b. Genetic study of longevity in Swedish Landrace sows. *Livestock Production Science* 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 hazard models. *J. Dairy Sci.* 85:1563-1577.

Youssef Y.M.K., Khalil M.H., Afifi E.A., El-Raffa A.M.E., Zaheds M., 2000. Heretability and non genetic factors for lifetime production traits in New Zealand white rabbits raised in intensive system of production. In: *Proc. of the 7th World Rabbit Congress, Valencia (Spain). Vol A, pp. 497 - 503.*

CHAPTER 2: OBJECTIVES

CHAPTER 2: OBJECTIVES

- 1.- Study how different environmental factors, production level and reproductive rhythm, influence the functional longevity in a meat rabbit population.

- 2.- Study the genetic determinism of the functional longevity in a meat rabbit population.

- 3.- Estimate the genetic and environmental correlations between functional longevity and prolificacy traits in a meat rabbit population.

- 4.- Constitute a maternal line of rabbits for meat production applying very high intensities of selection for longevity and culling levels near the average for prolificacy.

- 5.- Evaluate this new line, for longevity, prolificacy, cumulative production and kindling interval, in order to check the interest of the procedure of selection for hyper-longevity to create new lines.

CHAPTER 3: GENETIC ANALYSIS OF LONGEVITY IN RABBITS

CHAPTER 3: GENETIC ANALYSIS OF LONGEVITY IN RABBITS

3.1.- Analysis of factors influencing longevity of rabbit does⁵

3.1.1.- Abstract

A survival analysis, using a Cox proportional hazard model, was carried out to show the factors influencing longevity of the does in a selection nucleus. It was shown that the year-season, the combination between number of live born and reproductive rhythm, the combination between kindling order and reproductive rhythm and the age of the doe at first mating significantly affected doe longevity. Once the significant factors were known, the model was extended, including the sire effect and the maternal grandsire effect, to estimate the additive variance of the trait. This estimate was 0.088. Finally, to evaluate the effects of factors influencing longevity, an animal model was used, with the additive variance estimated previously. Inclusion of longevity as a selection criterion in rabbits does not seem advisable due to its low heritability; the trait is only recorded in females; the live females have the trait censored and its consideration makes the generation interval longer. An alternative could be developing new strains of rabbits by applying very high selection intensities in commercial populations, similar to the schemes successfully applied in pigs and rabbits.

Keywords: Cox models, rabbits, longevity, proportional hazard models.

3.1.2.- Introduction

Presently, improvement in reproductive traits such as litter size at different stages, mainly birth and weaning, and growth traits such as conversion index, indirectly evaluated by daily gain, are the principal objectives in rabbit breeding programs. (Baselga and Blasco, 1989; Rochambeau et al., 1994; García and Baselga, 2002).

Improvement of productive life of does could be interesting because the replacement rate is very high (120%), (Ramón and Rafel, 2002). The main causes of the end of the productive life of does are death and sickness during the first production

⁵ Published in *Livestock Production Science* with reference:
Sánchez J.P., Baselga M., Peiró R., Silvestre M.A., 2004. Analysis of factors influencing longevity of rabbit does 90:227-234

cycles, when the cost of the does has not been yet recovered. Due to the high annual replacement rate, the proportion of young does in production is high, and these does are more difficult to handle than older ones. All these factors show the importance of enhancing functional longevity in rabbits.

In prolific species, the length of productive life and related traits such as lifetime production have recently been treated by different methodologies in several studies. One peculiarity of longevity-related traits is that for some animals in study we have censored the data. This means that the true record for these animals, the length of its productive life, should be higher than the recorded one because culling was performed for reasons other than illness or because the animal was still alive when the analysis was carried out. Thus, the animal had not yet reached the event of interest (death or culling by illness) that allows to know the exact length of its productive life, but by not including them in the analysis (or including them as complete data) this would bias the results. Uncensored data are the opposite, which consists of data only from animals that have reached the event of interest. Youssef et al. (2000), using mixed linear models without taking into account the censored data, estimated by REML the heritabilities of the length of productive life and of cumulative production traits of rabbits. Several rabbit genetic types were compared for cumulative traits over one year and for survival throughout this period, using a linear model without censoring by Lukefahr and Hamilton (2000). Longevity and cumulative prolificacy were studied in a Landrace pig population by Bayesian methodology using linear models with censoring (Guo et al., 2001). Although linear models could take into account the censored data, they are not considered the best way to explain and analyse longevity and related traits because with these models time dependent factors cannot be used. On the other hand, with these methods the dependent variable is the time to failure, which is very asymmetric and generally has an unknown distribution.

Another approach more suitable to study longevity is survival methodology, which is based on analysing the hazard, which could have a known parametric distribution or which could be fitted by non parametric models. This method also allows the use of factors that change with time and the censored data are included naturally in the analysis. The use of this methodology has been limited in prolific mammals of interest in animal production although it has been applied in Landrace sows by Yazdi et al. (2000) and in rabbits by Garreau et al. (2001).

The aim of this work was to find the most important factors influencing doe longevity, using survival analysis methodology, (Kalbfleisch and Prentice, 1980), in a particular rabbit population, specially suited to the study of functional longevity, because in this population culling based on production traits is not carried out.

3.1.3.- Material and Methods

Data for this study were collected from the beginning of 1992 to the end of 2001 in a selection nucleus of rabbits, placed in Sant Carles de la Rápita (Tarragona, Spain). The animals belong to V strain (Estany et al., 1989). These animals were selected for litter size at 28 days after parturition and the genetic evaluation was carried out by a family index. In this nucleus, the management system is semi-intensive; the does are mated 11 days post-partum, and the young are weaned at 35 day. For a suitable genetic evaluation of animals in the nucleus, common culling criteria in commercial farms are not considered; i.e., does with low levels of production or no strong reproductive delay, are not culled.

The data set included the reproductive records of the 13 first generations, involving 2488 does. The pedigree file had 3031 records. Because of space limitations and the breeding program of the nucleus, 48% of the does were removed before the end of their productive lives. The remaining 52% were animals for which the date of death or culling as a result of various pathological problems was known. In Table 3.1.3.1 are presented the maximum, minimum and average times for censored and uncensored records.

Table 3.1.3.1.- Minimum, maximum and average productive life for censored and uncensored records in the data set.

	Censored ^a Records 1190 (47.83%)	Uncensored ^b Records 1298 (52.17%)
Minimum time (d)	25	1
Maximum time (d)	627	601
Average time (d)	397.80	189.46

^aRecords from does that had not completed their productive life.

^bRecords from does that had completed their productive life.

The individual record of each animal included the date of the first presentation to a male and the date of death or culling. The difference between these dates and the

censoring code were the response variables. Other factors that could affect as regression variables were collected:

- Year-Season (YS), time dependent variable, with 19 levels. The level changes every 6 months (1st January and 1st July, every year). This factor attempts to explain the effect acting over all the animals that are on the farm at the same time.
- The generation of the doe (GE), time independent variable.
- Kindling order (KO), time dependent, with 9 levels. The first one includes all the nulliparous does, the second includes all the primiparous does and so on until the ninth level, which includes the does with 8 or more parturitions.
- The interval between parturition and the next fertile mating. A time dependent variable with two levels: less than or more than 28 days in no-nulliparous does. In nulliparous does this factor measures the interval between the first presentation to male and the first fertile mating as greater or less than 21 days. This factor changes every 28 days post-partum (day of control) and shows the reproductive rhythm of the doe (RR). At t=0 (time of the first presentation to a male) all does are in level 1 of RR, but at t=21 they can change to level 2 if at this time they are not pregnant.
- Number born alive (BA), a time dependent variable with 8 levels. All the does are included in the first level when they are nulliparous; the second includes does that are not nulliparous and have zero or one live born, does with 2 or 3 born alive are included in the third level and so on until the eighth level, which includes the does with 11 or more born alive. This factor could change every parturition.
- Age of the doe at the first presentation to male (AFM), a time independent covariate.

The survival analysis methodology allows us to include censored data in the analysis of different factors that could influence the risk of death or culling of a doe. Proportional hazards models were used in this study, which are described by the following equation (Kalbfleisch and Prentice, 1980).

$$h_i(t | \vec{x}_i') = h_0(t) \exp\{\vec{x}_i' \vec{\beta}\}$$

Where:

$h_i(t | \bar{x}'_i)$ is the hazard of animal i at time t , affected by covariates indicated by \bar{x}'_i ; $h_0(t)$ is the baseline hazard at time t ; \bar{x}'_i is the vector which selects the level of the different factors that are affecting the animal i and $\bar{\beta}$ is the vector of regression coefficients for the different factors.

We used the survival methodology in a progressive way to:

1.- Check the proportional hazard assumption, using a graphical test of the Kaplan-Meier estimator of the survival function (a non-parametric estimator that uses the crude data), in order to determine if a stratified analysis per generation is needed or not. The generation factor is time independent and it allows the partition of the whole population of animals into different and separate subpopulations that could have distinct survival functions.

2.- Check if a Weibull distribution is suitable for defining the baseline hazard function or if a semiparametric Cox model is preferable, using a graphical test of the Kaplan-Meier estimator of the survival function. In proportional hazard models, the baseline hazard function could be defined by a parametric function (parametric models) or could be kept arbitrary (semiparametric models), in the sense that it keeps undefined through the use of a partial likelihood which does not depend on the baseline function (Cox Model), or is defined by a step function arbitrarily chosen, most of the times the jumps in this arbitrary step function are assumed to occur at the observed failure times (Kalbfleisch and Prentice, 1980).

3.- Determine the fixed factors that are significant to explain longevity, using the likelihood ratio test (Kalbfleisch and Prentice, 1980).

4.- Extend the model of significant fixed effects, including sire and maternal grandsire effects, in order to estimate the variance associated with the sire effect. (Ducrocq and Casella, 1996). This value was used to estimate the heritability on the logarithm of the integrated hazard function (including only time dependent covariates) scale (Korsgaard et al., 2002), using the following equation.

$$h^2 = \frac{4\sigma_s^2}{\sigma_s^2 + 1/4\sigma_s^2 + \pi^2 / 6}$$

Where σ_s^2 is the estimated sire variance and $\pi^2 / 6$ is the variance of the extreme value distribution.

This heritability, although theoretically correct in this scale, is not adequate to compute reliabilities of the genetic evaluations and responses to selection in the original scale. Korsgaard et al. (1999) proved that if a linear approximation from this scale to the original scale is acceptable, the same heritability is valid in both scales. Ducrocq (1999) and Yazdi et al. (2002) have addressed the problem of computing a reasonable heritability value to allow computation of reliabilities and responses to selection in sire models under the Weibull distribution of the hazard.

5.- Estimate the effects of the fixed factors through an animal model in agreement with the heritability estimated with the sire-maternal grandsire model.

All the analyses were carried out using the survival kit 3.0 (Ducrocq and Sölkner, 1998).

3.1.4.- Results and Discussion

3.1.4.1.- Checking the proportional hazards assumption

In order to check the necessity of a stratified analysis by generation, a plot of $\log(-\log(S_{KM,n}(t)))$ versus $\log(\text{time})$ was carried out. The $S_{KM,n}(t)$ is the Kaplan-Meier estimator of the survival function for generation n . Straight parallel lines would indicate that the baseline hazard functions are proportional for any pair of strata (generations). This plot is presented in figure 3.1.4.1.1.

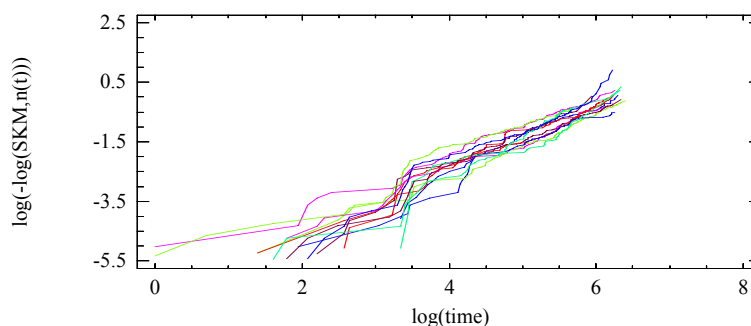


Figure 3.1.4.1.1.- Graphical test of the proportional hazard assumption between different generations. $S_{KM,n}(t)$, estimator of Kaplan-Meier of the survival function for generation n ; time is days since first presentation to a male.

In figure 3.1.4.1.1, the parallelism occurs throughout the time. In the first days of production the lines are not parallel; this may be due to the low number of animals (around 200) in each stratum. Similar results were obtained by Ducrocq et al. (2000) in

hens, but they obtained a better parallelism with a higher number of animals per stratum.

3.1.4.2.- Checking the suitability of a Weibull baseline hazard function

In order to check the Weibull hazard function as the baseline hazard function, the same graphical test is carried out, but for the whole population: $\log(-\log(S_{KM}(t)))$ versus $\log(\text{time})$. This plot should display a straight line. The results of this test are shown in figure 3.1.4.2.1. The fit fails at around 30 days. This could be due to a high proportion of deaths (15.72%) in the first parturition. It was therefore decided to use a semiparametric Cox model. Yazdi et al. (2000) studied the length of productive lives in pigs from the first parturition and they had a similar plot, but they decided to use the Weibull model, because this model is computationally less demanding than the Cox model and the size of their experiment was too large to apply the latter model. In our work the number of animals was low and a semiparametric analysis was feasible.

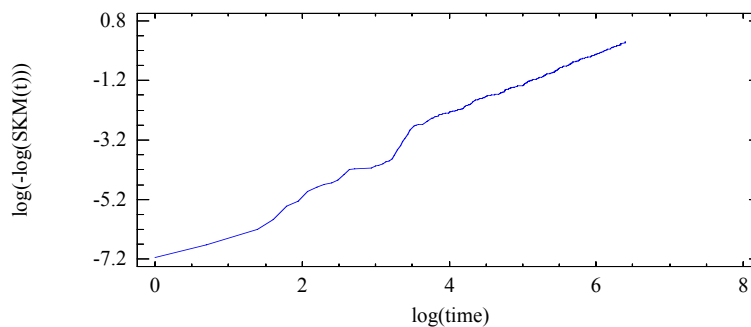


Figure 3.1.4.2.1.- Graphical test of the assumption of a Weibull distribution for the baseline hazard function. $S_{KM}(t)$, estimator of Kaplan-Meier of the survival function for the whole population; time is days since first presentation to a male.

3.1.4.3.- Defining the model

The likelihood ratio tests between successive nested models are shown in table 3.1.4.3.1. Factors are included one by one in the model. In this analysis, the generation effect is not included because it is confounded with the YS effect. Among the studied effects, only the interaction KO*BA had no significant effect. Thus, the final fixed effects model proposed was:

$$h_i(t | \vec{x}'(t)_i) = h_0(t) \exp \left\{ \vec{x}'(t)_{RR*BA_i} \vec{\beta}_{RR*BA} + \vec{x}'(t)_{YS_i} \vec{\beta}_{YS} + \vec{x}'(t)_{KO*RR_i} \vec{\beta}_{KO*RR} + AFM_i \beta_{AFM} \right\}$$

Where:

$h_i(t | \vec{x}'(t)_i)$ is the hazard of animal i at time t , affected by covariates indicated by $\vec{x}'_i(t) = \left\{ \vec{x}'(t)_{RR*BA_i}, \vec{x}'(t)_{YS_i}, \vec{x}'(t)_{KO*RR_i}, AFM_i \right\}$; $h_0(t)$ is the baseline hazard at time t , (in the Cox model is not defined by any function); $\vec{x}'(t)_{RR*BA_i}$ is the vector which selects the level of RR*BA combination which at time t is affecting animal i ; $\vec{x}'(t)_{YS_i}$ is the vector which selects the level of YS which at time t is affecting animal i ; $\vec{x}'(t)_{KO*RR_i}$ is the vector which selects the level of KO*RR combination which at time t is affecting animal i ; AFM_i is the age of the doe i at first matting; and $\vec{\beta}_{RR*BA}$, $\vec{\beta}_{YS}$, $\vec{\beta}_{KO*RR}$ and β_{AFM} are the corresponding vectors of regression coefficients for the respective factors.

Table 3.1.4.3.1.- Likelihood ratio test, including factors sequentially added to the fixed effects model

Factor ^a	TOTAL DF	-2 log (Like.)	CHI ²	Δ DF	p-value
NONE	0	19045.90			
YS	18	18963.97	81.93	18	< 0.001
KO	26	18599.97	364.00	8	< 0.001
RR	27	18456.92	143.05	1	< 0.001
KO*RR	35	18415.82	41.10	8	< 0.001
BA	41	18121.30	294.51	6	< 0.001
KO*BA	83	18073.84	47.46	42	0.260
RR*BA	89	18041.35	32.48	6	< 0.001
AFM	90	18036.74	4.61	1	0.032

^a Year-Season (YS), Kindling order (KO), Reproductive rhythm (RR), Interaction between KO and RR (KO*RR), Number of born alive (BA), Interaction between KO and BA (KO*BA), Interaction between RR and BA (RR*BA), Age at first presentation to male (AFM, d)

3.1.4.4.- Variance components estimation

In order to estimate the additive variance of the length of productive life of the does, the fixed effects model defined in 3.1.4.3 was extended to include the random effects of sire and the maternal grand sire. The model was:

$$h_i(t | \vec{x}'(t)_i) = h_0(t) \exp \left\{ \vec{x}'(t)_{RR*BA_i} \vec{\beta}_{RR*BA} + \vec{x}'(t)_{YS_i} \vec{\beta}_{YS} + \vec{x}'(t)_{KO*RR_i} \vec{\beta}_{KO*RR} + AFM_i \beta_{AFM} + \vec{z}'_{i,s} \vec{s} + 1/2 \vec{z}'_{i,mgs} \vec{s} \right\}$$

Where the new terms are:

\bar{s} is the vector of additive effects of the sires; $\bar{z}'_{i,s}$ is the vector which selects the sire of the doe i in \bar{s} , and $\bar{z}'_{i,mgs}$ is the vector which selects the maternal grandsire of the doe i in \bar{s} .

The mode, the mean, the standard deviation and the skewness coefficient of the marginal posterior distribution of the sire variance were 0.022, 0.025, 0.009 and 0.615 respectively. The estimated heritability, using the mode of the posterior distribution of the sire variance, was 0.053.

There are several studies in which the heritability for length of productive life of rabbits is estimated, but because the methods and the scales are different, their results are not comparable with ours. These estimates are somewhat higher than ours, 0.13 in Youssef et al. (2000) and 0.05, 0.10 or 0.24, depending on the model, in Garreau et al. (2001).

3.1.4.5.- Effects estimation

In order to estimate the effects of the factors, an animal model was proposed and the additive variance estimated in 3.1.4.4 was used in this analysis. The model was:

$$h_i(t | \bar{x}'(t)_i) = h_0(t) \exp \left\{ \bar{x}'(t)_{RR*BA_i} \bar{\beta}_{RR*BA} + \bar{x}'(t)_{YS_i} \bar{\beta}_{YS} + \bar{x}'(t)_{KO*RR_i} \bar{\beta}_{KO*RR} + AFM_i \beta_{AFM} + \bar{z}'_i \bar{u} \right\}$$

Where the new terms are:

\bar{u} is the vector of additive values and \bar{z}'_i is the vector that selects the animal effect in \bar{u} .

The relative (to the 19th level) risks for the YS effect, are shown in figure 3.1.4.5.1. The latter YS levels had higher relative risk than the earlier ones. This may be explained by a higher disease incidence in the latter generations. The maximum risk ratio was between the second half of 1994 (YS₆) and the first half of 1998 (YS₁₃). This means that culling of a doe that lived in the YS₁₃ was 3.11 times more likely than a doe that lived in YS₆, with the other factors at the same level.

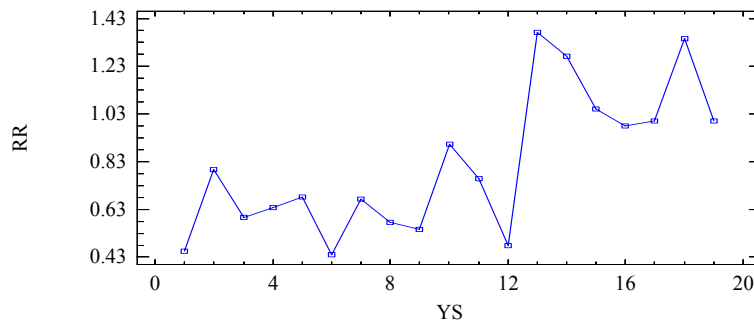


Figure 3.1.4.5.1.- RR, relative risks for the year-season (YS) effect.

For the different levels of the combination between RR and BA the hazard coefficients, the standard errors, the relative risks and the number of uncensored records are shown in table 3.1.4.5.1. Apart from the first level of BA referring to nulliparous does, the second level (which included the does with zero or one born alive in any kindling) had the greatest relative risk for both RR. As the number of live born increased, the relative risk became lower, but always non-pregnant does 28 days after kindling (RR=2) had a greater risk of culling. Similar results for the factor number born alive were obtained by Garreau et al. (2001). In our data set there was no culling for productive reasons, so low levels of BA would be indicators of sick does, because they had the highest relative risks.

Table 3.1.4.5.1.- Estimates within reproductive rhythm (RR) of the hazard coefficient (Coef.), standard error (S.E.), relative risk (rr) and number of uncensored data (n_{unc}) for each level of born alive (BA).

BA	RR 1				RR 2			
	Coef.	S.E.	rr	n_{unc}	Coef.	S.E.	rr	n_{unc}
1	-0.81	0.41	0.442	141	-0.37	0.34	0.687	43
2	-3.78	0.49	0.023	137	-3.81	0.58	0.022	15
3	-5.15	0.51	0.006	41	-4.47	0.59	0.011	11
4	-5.49	0.50	0.004	60	-4.53	0.56	0.018	17
5	-5.80	0.50	0.003	82	-4.39	0.54	0.012	29
6	-5.88	0.49	0.003	139	-4.38	0.52	0.012	67
7	-5.89	0.49	0.003	181	-4.48	0.52	0.011	83
8	-5.89	0.49	0.003	192	-4.95	0.53	0.007	60

To study the effects of the combination between KO and RR simple consideration of the relative risks between the levels of KO*RR makes no sense because KO progresses with time. It is thus necessary to take into account the

corresponding change of time in the baseline hazard function (Gröhn et al., 1997). Consequently, in order to illustrate the KO*RR effects on the length of productive life of the does, the survival function of does, in the same level of BA (0 and 1 born alive) and YS levels (3, 4 and 5) but with different reproductive patterns are shown in Figure 3.1.4.5.2. Does with delayed first mating (D,E) had a lower survival probability at the first parity ($S(59)=0.7$) than those that became pregnant in the first 21 days of productive life (A,B,C; $S(31)=0.96$). This was because they were older and in a different physiological state, but if the comparison is done at the same time, $t=59$, the former had a survival probability 12 % lower than the latter. To study the effect of delay in the second parturition, records from does with a delay in the first parturition (D vs E) or without this delay (A and B vs D) were used. In the first case, does delayed for the second parity (E) kindled at 122 d, had a $S(122)=0.32$. The does without delay had the second parity at 101 d, and their $S(101)$ was 0.44. Here, the most interesting fact is that at the same time, 101 d, the survival of the delayed does was 15 % greater than the survival of no-delayed does. This could easily be explained because at this stage the does are still growing and the kindling would be an important risk factor. If the comparison is done at $t=122$ there were no differences in survival. The same fact is observed when we compare does without delay in the first parity, but for these does the differences at the time of the second parturition with non-delayed does was only 5 %, in favour of the delayed does. To illustrate the effect of the delay in latter kindlings, if we compare A vs B does, this delay had a very low negative effect on survival. However, does without delay in the last cycles, but with delay at the beginning of their career, always had lower survival probabilities. Thus, delays at the beginning of productive life had negative effects throughout the life of the animal.

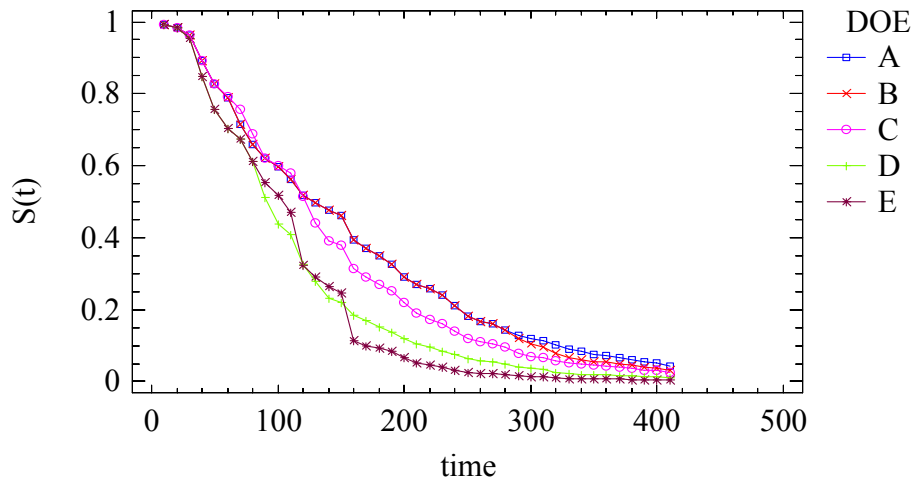


Figure 3.1.4.5.2.- Survival function of does^a with different reproductive patterns.

^aA: Doe that never has a delay greater than 28 days.

B: Doe that has a delay greater than 28 days in the 6th cycle.

C: Doe that has a delay greater than 28 days in 2nd cycle.

D: Doe that has a delay greater than 21 days in the 1st cycle.

E: Doe that has a delay greater than 21 days in the 1st cycle and greater than 28 days in 2nd cycle.

Age at first mating was treated as a covariate; the mean was 148 d and the standard deviation 14 d. The estimated regression coefficient was -0.008 (0.003). This means that the relative risk between 120 and 176 days old does (mean \pm two standard deviations) was 1.56. Thus, older the does at first mating lower the risk of culling.

3.1.5.- Conclusions and implications

The most important factors influencing longevity in does were the combination between reproductive rhythm and number born alive (RR*BA) and the combination between kindling order and reproductive rhythm (KO*RR). The year season effect showed an erratic behaviour that could be associated to periods with different incidence of diseases in the nucleus. Age of the does at first mating showed a significant but not very important effect.

The longevity in our population showed low additive variance and consequently low heritability. Improving longevity in rabbit populations is highly desirable, in order to reduce replacement rates. However, inclusion of longevity as a selection criterion in rabbits does not seem advisable due to its low heritability; the trait is only recorded in

females; the live females have the trait censored and its consideration makes the generation interval longer. One possibility to improve longevity would be to develop new strains of rabbits by applying very high selection intensities in commercial populations, similar to the schemes successfully applied in pigs and rabbits. (Bichard and Seidel, 1982; Sorensen and Vernersen, 1991; Noguera et al., 1997; Cifre et al., 1998). The criterion would be to select does with very long productive lives and average litter sizes above the mean of the commercial population.

3.1.6.- Acknowledgements

The authors thank Dr. Vincent Ducrocq for his useful advice and Mr. Neil Macowan for revising the English version of the manuscript. This work was supported by the Spanish Research Project AGL2000-0595-C03 (CYCIT).

3.1.7.- References

- Baselga, M., Blasco, A., 1989. Mejora genética del conejo de producción de carne. Agroguías mundi-prensa. Madrid.
- Bichard, M., Seidel, C.M., 1982. Selection for reproductive performance in maternal lines of pigs. In: Proceedings of 2nd World Congress on Genetics Applied to Livestock Production, Vol. VIII, Madrid, Spain, pp. 565-569.
- Cifre, J., Baselga, M., García-Ximénez, F., Vicente, J.S., 1998. Performance of a hyperprolific rabbit line. I. Litter size traits. *J. Anim. Breed. Genet.* 115 , 131-138.
- 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 V3.0, a package for large analyses of survival data. In: Proceedings of 6th World Congress on Genetics Applied to Livestock Production, Vol. 27, Armidale, Australia, pp. 447- 450.
- Ducrocq, V., 1999. Topics that may deserve further attention in survival analysis applied to dairy cattle breeding – some suggestions. In: Proceeding of

International Workshop on Genetics Improvement of Functional Traits in Cattle (GIFT); Longevity. Jouy-en-Josas, France. Bulletin No 21, pp. 181 – 189.

Ducrocq, V. , Besbes, B., Protais, M., 2000. Genetic improvement of laying hens viability using survival analysis. *Genet. Sel. Evol.* 32, 23-40.

Estany, J., Baselga, M., Blasco, A., Camacho, J., 1989. Mixed model methodology for the estimation of genetic response to selection in litter size of rabbits. *Livest. Prod. Sci.* 21, 67-75.

García, M. L., Baselga, M., 2002. Estimation of genetic response to selection in litter size of rabbits using a cryopreserved control population. *Livest. Prod. Sci.* 74, 45 – 53.

Garreau, H., Larzul, C., Ducrocq, V., 2001. Analyse de longévité de la souche de lapins INRA 1077. In: 9èmes Journ. Rech. Cunicole, Paris, France, pp. 217 - 220.

Gröhn, Y. T., Ducrocq, V., Hertl, J. A., 1997. Modelling the effect of a disease on culling: An illustration of the use of time-dependent covariates for survival analysis. *J. Dairy Sci.* 80, 1755 – 1766.

Guo, S., Gianola, D., Rekaya, R., Short, T., 2001. Bayesian analysis of lifetime performance and prolificacy in Landrace sows using a linear mixed model with censoring. *Livest. Prod. Sci.* 72, 243 – 252.

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

Korsgaard, I. R., Andersen, A.H., Jensen, J., 1999. Discussion of heritability of survival traits. In: *Proceeding of International Workshop on Genetic Improvement of Functional Traits in Cattle (GIFT): Longevity, Jouy-en-Josas, France. Bulletin No 21, pp. 31 – 35.*

- Koorsgaard, 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.
- Lukefahr, S. D., Hamilton H.H., 2000. Longevity and cumulative litter productivity in straightbred and crossbred Californian and New Zealand White does, In: Blasco, A. (Ed.), *Proceeding of 7th World Rabbit Congress*, Vol. A, Valencia, Spain, pp. 463 – 468.
- Noguera, J.L., Alfonso, L., Babot, D., Pérez – Enciso, M., Estany, J., 1997. Resultados de un experimento de selección de tamaño de camada mediante un esquema hiperprolífico en porcino. *ITEA, VII Jornadas sobre Producción Animal*, Vol Extra 18 (1), pp. 391-393.
- de Rochambeau, H., Bolet, G., Tudela, F., 1994. Long-term selection. Comparison of two rabbit strains. In: Smith, C., Gavora, J.S., Benkel, B., Chesnais, J., Fairfull, W., Gibson, J.P., Kennedy, B.W., Brunside, E.B. (Eds.), *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*, Vol. 19., Guelph, Canada, pp. 257-260.
- Ramón, J., Rafel, O., 2002. 1991-2000. Diez años de gestión global en España. In: *Expoaviga 2002, X jornada cunícola*, Barcelona, pp.113–117.
- Sorensen, D. A., Vernersen, A.H., 1991. Large scale selection for number of born piglets using an animal model. In: 42nd Annual meeting of the European Association of Animal Production. Commission on Animal Genetics, Session 3. Berlin.
- Yazdi, M. H., Rydhmer, L., Ringmar-Cederberg, E., Lundeheim, N., Johansson, K., 2000. Genetic study of longevity 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 hazard models. *J. Dairy Sci.* 85:1563-1577.

Youssef, Y.M.K., Khalil, M.H., Afifi, E. A. , El – Raffa, A.M.E., Zaheds, M., 2000. Heritability and non genetic factors for lifetime production traits in New Zealand White rabbits raised in intensive system of production. In: Blasco, A. (Ed.), *Proceeding of 7th World Rabbit Congress, Vol. A, Valencia, Spain, pp. 497 - 503*

3.2.- Analysis of rabbit does longevity using a semiparametric log-Normal animal frailty model with time-dependent covariates⁶

3.2.1.-Abstract

Data on doe longevity in a rabbit population were analyzed using a semiparametric log-Normal animal frailty model. Longevity was defined as time from the first positive pregnancy test to death or culling due to pathological problems. Does culled for other reasons had right censored records of longevity. The model included time dependent covariates associated with year by season, with the interaction between physiological state and the number of young born alive, and between order of positive pregnancy test and physiological state. The model also included an additive genetic effect and a residual in log frailty. Properties of marginal posterior distributions of specific parameters were inferred from a full Bayesian analysis using Gibbs sampling. All of the fully conditional posterior distributions defining a Gibbs sampler were easy to sample from, either directly or using adaptive rejection sampling. The marginal posterior mean estimates of the additive genetic variance and of the residual variance in log frailty were 0.247 and 0.690.

Key words: animal model, longevity in rabbit does, semiparametric log-Normal frailty model, survival analysis, time dependent covariates

3.2.2.- Introduction

In rabbit does for meat production the average annual replacement rate is around 120 % (Ramón and Rafel; 2002). Moreover, around 50 % of the dead and culled does are in one of the first three kindlings (Rosell, 2003). This implies that the possibilities for culling due to low production are limited and that a high proportion of does are dead or culled before the investment in buying them is recovered. It is therefore, desirable to increase longevity of rabbit does and one possibility to achieve this objective is through out directional selection. In order to improve longevity through selective breeding, it is necessary to demonstrate that additive genetic variation exists.

⁶ Submitted to Genetics Selection and Evolution:
Sánchez J.P., Korsgaard I.R., Damgaard L.H., Baselga, M., 2005. Analysis of rabbit does longevity using a semiparametric log-Normal animal frailty model with time-dependent covariates

In rabbits one male is mated with a few does, the proportion of full-sibs in the data is high and females are closely related. In populations with this family structure the assumptions underlying sire models are far from being satisfied, and to use animal models should be preferred. Another characteristic of rabbits is the very high reproductive rhythm, which implies that a female change from one physiological state (e.g. lactating, pregnant, empty) to another within just a few days. The risk of culling/death is expected to differ in different physiological states, therefore it is desirable to use an animal model with time dependent covariates for analysing longevity records of rabbits.

Apart from a study of Diaz et al. (2002) where they used the Survival Kit (Ducrocq and Sölkner, 1998) to estimate the additive genetic variance for longevity of beef cattle using an animal model, we are not aware of any other applications of the Survival Kit in which an animal model has been applied. In a simulation study Ducrocq and Casella (1996) analysed data simulated according to a balanced halfsib design with an animal model using the Survival Kit and found that the estimated sire component of the variance was upward biased. They argued that this bias may be due to limitations of the Laplacian integration used to obtain the marginal posterior distribution of the genetic variance component, because of the high number of parameters to integrate out in animal models and because of the lack of relationships on the female side. It is likely that these simulation results, in combination with the larger computation time in animal models compared to sire models, have lead to the domination of the use of sire models over animal models in genetic applications (Sánchez et al., 2004; Garreau et al., 2001; Yazdi et al., 2000). Another reason may be that survival models have mainly been used to address genetic aspects of longevity of dairy cows in which the family structure assembles that of a halfsib design to a higher extent than that of a rabbit population.

The semiparametric animal frailty model used by Korsgaard et al. (1998) to analyse records of time until first occurrence of a respiratory disease of beef bulls is a generalization of the corresponding model in Survival Kit in that it includes a normally distributed residual effect on the log-frailty scale. This model recognises that individual heterogeneity, due to environmental factors, generally exists. Korsgaard et al. (1998) derived and implemented the animal model without time-dependent covariates and in a Bayesian setting using Gibbs sampling, from which the marginal posterior distribution of the additive genetic variance and the residual variance in log-frailty were easily obtained.

The purpose of this paper was to infer additive genetic and environmental aspects of length of productive life of rabbit does using an animal model. To achieve this objective we extended the model proposed by Korsgaard et al. (1998) to include time dependent covariates.

3.2.3.-Material and Methods

3.2.3.1.-Data

Data on rabbit does longevity were collected from the beginning of 1992 to the end of 2001 in a selection nucleus of rabbits, located in Sant Carles de la Rápita (Tarragona, Spain). The animals belong to the V strain (Estany et al., 1989), where animals were selected for litter size, 28 days after parturition, based on a family index. In this nucleus, the does are mated 11 days postpartum, the pregnancy test is carried out by abdominal palpation 12 days after mating and the young are weaned at 35 days of age.

The data set included reproductive records of the first 13 generations, involving 2,400 does. The pedigree file included a total of 3,031 animals. One male was on average mated with 2.2 females having 7.7 offspring, and one female with 1.03 males having 3.6 offspring. Because of space limitations and the breeding program of the nucleus, 50% of the does were removed before the end of their productive life - these animals have censored records of longevity. The remaining 50% of the does either died or were culled because of pathological problems - these animals have uncensored records of longevity. Animals still alive at day 586 (maximum uncensored record of longevity) were censored at age 586. Summary statistics for censored and uncensored records are presented in Table I.

Table I. Minimum (Min), Maximum (Max), and Average (Mean) values (in days) among censored and uncensored records of rabbit does longevity.

	Censored	Uncensored
Min	13	1
Max	586	586
Mean	377.9	188.2

Records of individual animals included the date of first positive pregnancy test and the date of death or culling. The response variables were the time between these

dates and the censoring code. Individual records also included dates of the different positive pregnancy tests, the number of young born alive in every parturition, and the length of the different physiological states (Pregnant, Lactating, Empty and Pregnant&Lactating).

3.2.3.2.-Methods

The semiparametric log-Normal animal frailty model (Andersen et al., 2000) was used to analyse rabbit doe longevity. Model parameters were inferred from a fully Bayesian analysis using Gibbs sampling (Geman and Geman, 1984). After having specified the conditional likelihood and the prior distribution, a set of fully conditional posterior distributions defining a Gibbs sampler was derived. Some of the fully conditional posterior distributions are well known distributions that are easy to sample from, the remaining can all be shown to be log-concave and Adaptive Rejection Sampling (ARS) was used to obtain samples from these distributions (Gilks and Wild, 1992).

3.2.3.2.1.-Model

Let T_i and C_i be random variables representing the survival and the censoring time of animal i . The data on animal i are (y_i, δ_i) , where y_i is the observed value of $Y_i = \min(T_i, C_i)$ and δ is an indicator random variable, equal to 1 if $T_i \leq C_i$, and 0 if $T_i > C_i$. Vectors $\mathbf{y}, \boldsymbol{\delta}$ will denote data on all animals. In the semiparametric frailty model (Andersen et al., 1992) it is assumed that the hazard function of T_i conditional on random effects, \mathbf{a} and \mathbf{e} , and model parameters $\boldsymbol{\theta} = (\lambda_0(\cdot), \boldsymbol{\beta}, \sigma_a^2, \sigma_e^2)$ is given by

$$\lambda_i(t|\boldsymbol{\theta}, \mathbf{a}, \mathbf{e}) = \lambda_0(t) \exp\{\mathbf{x}'_i(t)\boldsymbol{\beta} + a_i + e_i\}, \quad (1.1)$$

where $\lambda_0(t)$ is the baseline hazard function, which is approximated by a piecewise constant function given by $\lambda_0(t) = \lambda_{0_m}$ for $t \in (\tau_{m-1}, \tau_m]$; $m = 1, \dots, M$, where τ_1, \dots, τ_M are the M different ordered survival times, $0 < \tau_1 < \dots < \tau_M < \infty$; $\tau_0 = 0$ and $\tau_{M+1} = \infty$. $\mathbf{x}'_i(t)$ is the time dependent design vector of animal i , which is assumed to be a piecewise constant function of time, and $\boldsymbol{\beta}$ is the corresponding vector of regression

parameters. In this study all of the elements of $\mathbf{x}'_i(t)$ are either zeros or ones. $\mathbf{a} = (a_i)_{i=1, \dots, N}$ is a vector of additive genetic values, $\mathbf{a} | \boldsymbol{\theta} \sim N_N(0, A\sigma_a^2)$, where N is the number of animals in the pedigree, and $\mathbf{e} = (e_i)_{i=1, \dots, n}$ is a vector of residuals in log frailty, $\mathbf{e} | \boldsymbol{\theta} \sim N_n(0, I\sigma_e^2)$, where n is the number of animals with record. \mathbf{a} and \mathbf{e} are assumed to be independent given $\boldsymbol{\theta}$. In this model $\exp\{a_i + e_i\}$ is the frailty term.

In this study $\boldsymbol{\beta}$ included effects of the following covariates:

- A time dependent Year-Season effect (YS) with 37 levels. The level changes every 3 months (1st January, 1st April, 1st July and 1st October every year). This factor accounts for the effect acting over all contemporary animals on the farm.

- A time dependent interaction between order of positive pregnancy test (OPPT) (1st, 2nd and \geq 3rd) and physiological state (PS) (Pregnant, Lactating, Empty and Pregnant&Lactating). A doe was considered Pregnant after a positive pregnancy test, Lactating after parturition if at least one young was born alive, Empty if no young were born alive, or after end of lactation if the doe had a negative pregnancy test, and Pregnant&Lactating in the part of the lactation period where the doe was pregnant (tested positive for pregnancy). Note that by definition, females in first order of positive pregnancy test can not be included in the physiological state Pregnant&Lactating.

- A time dependent interaction between PS and number born alive (BA), where BA has 9 levels: The first level included does that were not nulliparous and had zero born alive, does with one or two born alive were included in the second level, does with three or four born alive were included in the third level and so on until the eighth level, which included the does with 13 or more born alive. All of the does were included in the ninth level of BA before first parturition. Note that, a female having 0 born alive at a particular parturition can not be included in the physiological states Lactating or Pregnant&Lactating in that kindling.

All these factors were included in the model because their effects were found to be significantly different from zero, based on the likelihood ratio tests carried out in the Survival Kit (Ducrocq and Sölkner, 1998) using a Cox model. In order to ensure identifiability of regression parameters ('fixed effects') in the model a set of constraints were imposed using the algorithm proposed by Korsgaard et al. (In preparation).

3.2.3.2.2.-Conditional Likelihood

In the following we let $\boldsymbol{\psi} = (\boldsymbol{\theta}, \mathbf{a}, \mathbf{e})$ denote the augmented vector of parameters. Under the assumption that conditional on $\boldsymbol{\psi}$ censoring is independent and noninformative of $\boldsymbol{\psi}$ (Andersen et al., 1992) the conditional likelihood can be written as

$$p(\mathbf{y}, \boldsymbol{\delta} | \boldsymbol{\psi}) \propto \left[\prod_{m=1}^M \lambda_{0_m}^{d(\tau_m)} \right] \times \left[\prod_{m=1}^M \prod_{i \in D(\tau_m)} z_i \exp\{\mathbf{x}'_i(t) \boldsymbol{\beta}\} \right] \times \exp \left\{ - \sum_{m=1}^M \lambda_{0_m} \left[\sum_{i \in R(\tau_m)} z_i \int_{\tau_{m-1}}^{\tau_m \wedge y_i} \exp\{\mathbf{x}'_i(s) \boldsymbol{\beta}\} ds \right] \right\}, \quad (1.2)$$

where $z_i = \exp\{a_i + e_i\}$, $D(\tau_m)$ is the set of animals that fail in $(\tau_{m-1}, \tau_m]$, (i.e. at time τ_m), $d(\tau_m)$ is the number of animals in $D(\tau_m)$ and $R(\tau_m)$ is the set of animals that are observed during the whole interval or part of the interval $(\tau_{m-1}, \tau_m]$. The upper limit of the integral sign $(\tau_m \wedge y_i)$ means minimum of τ_m or y_i .

3.2.3.2.3.-Prior specification

A priori β_b , $b = 1, \dots, P$; λ_{0_m} , $m = 1, \dots, M$; σ_a^2 and σ_e^2 are assumed to be mutually independent. The prior of β_b was assumed to be improper uniform, i.e. $p(\beta_b) \propto 1$ for $-\infty < \beta_b < \infty$ for $b = 1, \dots, P$. The priors of the variance components σ_a^2 and σ_e^2 were assumed to be improper uniform, i.e. $p(\sigma_a^2) \propto 1$ and $p(\sigma_e^2) \propto 1$ for $0 < \sigma_a^2, \sigma_e^2 < \infty$. *A priori* it was assumed that random variables associated with the piecewise constant steps of the baseline hazard, λ_{0_m} , were independent with $p(\lambda_{0_m}) \propto \lambda_{0_m}^{-1}$ for $0 < \lambda_{0_m} < \infty$ for $m = 1, \dots, M$.

3.2.3.2.4.-Fully conditional posterior distributions and implementation of the Gibbs Sampler

A Bayesian analysis using Gibbs sampling requires a set of starting values of augmented parameters, $\boldsymbol{\psi}$, and sampling from the set of fully conditional posterior distributions of all the elements of this augmented parameter vector, $\boldsymbol{\psi}$. We use single

updating for all the elements of $\boldsymbol{\psi}$. The fully conditional posterior distribution of every parameter is proportional to the posterior distribution of $\boldsymbol{\psi}$, which up to proportionality is given by the product of the conditional likelihood (1.2) and the prior of $\boldsymbol{\psi}$. The fully conditional posterior distributions of all the elements of $\boldsymbol{\psi}$ are specified in the following, where for a particular element, φ , of $\boldsymbol{\psi}$ we let $\boldsymbol{\psi}_{\setminus\varphi}$ denote all elements in $\boldsymbol{\psi}$ except φ . The starting values for all the coordinates of $\boldsymbol{\beta}$, \mathbf{a} and \mathbf{e} were set equal to zero and the initial values for the additive genetic and the residual variances were 0.2 and 0.3.

The fully conditional posterior distribution of λ_{0_m} , $m=1,\dots,M$, is up to proportionality given by

$$p\left(\lambda_{0_m} \mid \mathbf{y}, \boldsymbol{\delta}, \boldsymbol{\psi}_{\setminus\lambda_{0_m}}\right) \propto \lambda_{0_m}^{d(\tau_m)-1} \exp\left\{-\lambda_{0_m} \Omega_{\lambda_{0_m}}\right\},$$

where $\Omega_{\lambda_{0_m}} = \sum_{i \in R(\tau_m)} \exp\{a_i + e_i\} \int_{\tau_m-1}^{\tau_m \wedge y_i} \exp\{\mathbf{x}'_i(s)\boldsymbol{\beta}\} ds$. This can be recognized as a Gamma distribution with shape parameter $d(\tau_m)$ and scale parameter $\Omega_{\lambda_{0_m}}^{-1}$. However the value of λ_{0_m} is expected to be close to zero, the lower bound of the domain, and to avoid numerical instabilities, first γ_{0_m} was sampled from a log Gamma distribution with parameters $d(\tau_m)$ and $\Omega_{\lambda_{0_m}}^{-1}$ (using ARS), next we let $\lambda_{0_m} = \exp\{\gamma_{0_m}\}$, this is a realized value from the fully conditional posterior distribution of λ_{0_m} .

For an animal i with a record, the additive genetic value, a_i , was sampled using ARS from the fully conditional posterior distribution given by

$$p\left(a_i \mid \mathbf{y}, \boldsymbol{\delta}, \boldsymbol{\psi}_{\setminus a_i}\right) \propto \exp\left\{a_i \left(\delta_i - \frac{1}{\sigma_a^2} \sum_{\substack{j=1 \\ j \neq i}}^N A^{ij} a_j\right)\right\} \times \\ \exp\left\{-\frac{A^{ii}}{2\sigma_a^2} a_i^2\right\} \exp\left\{-\exp\{a_i\} \Omega_i^a\right\},$$

where A^{ij} is the (i, j) th element of the inverse of the additive genetic relationship matrix and $\Omega_i^a = \exp\{e_i\} \int_0^{y_i} \lambda_0(s) \exp\{\mathbf{x}'_i(s)\boldsymbol{\beta}\} ds$. For an animal i without a record the

additive value was sampled from a Normal distribution with mean $-\frac{1}{A^{ii}} \sum_{\substack{j=1 \\ j \neq i}}^N A^{ij} a_j$ and variance $\frac{\sigma_a^2}{A^{ii}}$.

The additive variance was sampled from the Inverted Gamma distribution with parameters $\frac{N}{2}-1$ and $\left(\frac{\mathbf{a}'\mathbf{A}\mathbf{a}}{2}\right)^{-1}$, i.e. the fully conditional posterior density of σ_a^2 is up to proportionality given by $(\sigma_a^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{\sigma_a^2} \left(\frac{\mathbf{a}'\mathbf{A}\mathbf{a}}{2}\right)\right\}$.

The b 'th regression parameter, β_b ; $b=1, \dots, P$, was sampled, using ARS, from the fully conditional posterior distribution of β_b , with density up to proportionality given by

$$p(\beta_b | \mathbf{y}, \boldsymbol{\delta}, \boldsymbol{\Psi}_{\setminus \beta_b}) \propto \exp\{\beta_b d(\beta_b)\} \exp\{-\exp\{\beta_b\} \Omega_b^\beta\}$$

where $d(\beta_b)$ is the number of animals that died affected by level b of the covariates (i.e. the number of animals that died with $x_{ib}(y_i)=1$) and

$$\Omega_b^\beta = \sum_{i=1}^n \exp\{a_i + e_i\} \int_0^{y_i} \lambda_0(s) I\{x_{ib}(y_i)=1\} \exp\left\{\sum_{\substack{p=1 \\ p \neq b}}^P x_{ip}(s) \beta_p\right\} ds. \text{ In } \Omega_b^\beta, I\{x_{ib}(y_i)=1\}$$

is the indicator function which takes the value 1 if the b 'th coordinate of the design vector of animal i at time s , $x'_{ib}(s)$, is equal to 1 and 0 otherwise.

The residual value e_i , for $i=1, \dots, n$, was sampled, using ARS, from the fully conditional posterior distribution given by

$$p(e_i | \mathbf{y}, \boldsymbol{\delta}, \boldsymbol{\Psi}_{\setminus e_i}) \propto \exp\{e_i \delta_i\} \exp\left\{-\frac{1}{2\sigma_e^2} e_i^2\right\} \exp\{-\exp\{e_i\} \Omega_i^e\},$$

where $\Omega_i^e = \exp\{a_i\} \int_0^{y_i} \lambda_0(s) \exp\{\mathbf{x}'_i(s) \boldsymbol{\beta}\} ds$.

The residual variance was sampled from the Inverted Gamma distribution with parameters $\frac{n}{2}-1$ and $\left(\frac{\mathbf{e}'\mathbf{e}}{2}\right)^{-1}$, i.e. the fully conditional posterior density of σ_e^2 is up to proportionality given by $(\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{\sigma_e^2} \left(\frac{\mathbf{e}'\mathbf{e}}{2}\right)\right\}$.

In the model with time dependent covariates, all of the fully conditional posterior distributions of elements of $\boldsymbol{\psi}$ are very similar to those obtained by Korsgaard et al. (1998) in the model without time dependent covariates. The main difference is that the integral terms in the Ω 's become more involved and computationally more time consuming in models with time dependent covariates.

This Gibbs sampler was implemented in a Fortran 90 program, in which we used the subroutines by Gilks and Wild (1993) for adaptive rejection sampling.

After 100,000 rounds of the Gibbs sampler, 9,000 samples of model parameters were saved with a sampling interval of 100; so the total length of the chain was 1,000,000 rounds.

3.2.4.-Results

3.2.4.1.-Variances components

Table 3.2.4.1.1 shows marginal posterior summary statistics for the variance components and for the heritabilities on the log frailty scale ($h_{nor}^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$) and

on the $\log(g_i(t))$ scale ($h_{log}^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2 + \frac{\pi^2}{6})$) where $g_i(t)$ is given by

$$\int_0^t \lambda_0(s) \exp\{\mathbf{x}'_i(s)\boldsymbol{\beta}\} ds \quad (\text{Korsgaard et al., 2002}).$$

Note that, according to Korsgaard et al. (2002) $\log(g_i(T_i)) = -a_i - e_i + \varepsilon_i$, where ε_i follows an extreme value distribution

with $Var(\varepsilon_i) = \frac{\pi^2}{6}$, and with ε_i being independent of a_i and e_i .

Table 3.2.4.1.1.- Posterior mean (PM), posterior standard deviation (PSD), smallest interval with 95 % of the density (95% highest posterior density region) (95-SI), effective sample size (ESS) and lag 1 and lag 10 autocorrelations for the chains of the variance components σ_a^2 and σ_e^2 , and the heritabilities h_{nor}^2 and h_{log}^2 .

Parameter ^a	PM	PSD	95-SI	ESS	Lag 1	Lag 10
σ_a^2	0.247	0.115	0.046-0.482	717.8	0.784	0.203
σ_e^2	0.690	0.372	0.003-1.363	925.9	0.734	0.169
h_{nor}^2	0.300	0.163	0.017-0.623	464.8	0.847	0.319
h_{log}^2	0.095	0.04	0.018-0.174	787.3	0.783	0.205

^a σ_a^2 = additive genetic variance; σ_e^2 = residual variance; h_{nor}^2 = heritability on the log frailty scale; h_{log}^2 = heritability on the $\log(g_i(t))$ scale.

For the heritability on the $\log(g_i(t))$ scale the probability of a value greater than 0.01 was 99.34 %.

The variance components and the two different heritabilities (variance ratios) show good mixing properties according to the effective sample sizes obtained for the chains for these parameters. Because h_{nor}^2 and h_{log}^2 are functions of the variance components, these parameters are expected to show poor mixing properties. For these parameters convergence was assessed by visual inspection of the trace plot of the chains (not shown) and by Geweke's test (Geweke, 1992). The Z-score from this test was -1.18 for h_{nor}^2 and 1.23 for h_{log}^2 these values indicate that the test was passed satisfactorily since the Z-scores are within the interval (-1.96,1.96).

The marginal posterior distributions of h_{nor}^2 and h_{log}^2 are both right skewed implying that the mode is lower than the mean (Figure 3.2.4.1.1).

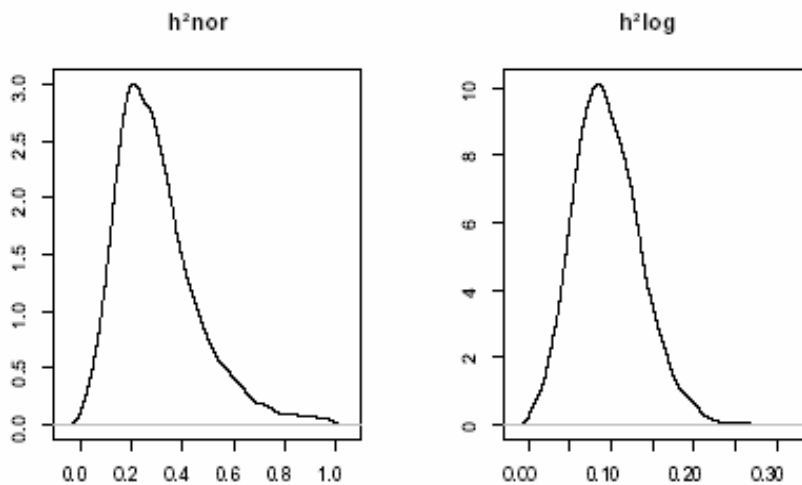


Figure 3.2.4.1.1.- Marginal posterior distributions for h_{nor}^2 and h_{log}^2 .

3.2.4.2.-Regression parameters

Based on the sampled values of β , estimable functions (γ) for the different levels of all the factors were computed in an similar way to how the least square means are computed in linear models. After the computation of these estimable functions the risk for each one of these γ_p was calculated as $\exp\{\gamma_p\}$. Note that the ratio between the risks of two different levels will be the relative risk between these two particular levels.

The posterior mean and the posterior standard deviation for the risks associated with the different levels of OPPT*PS are given in Table 3.2.4.2.1.

Table 3.2.4.2.1.- Posterior means (posterior standard deviations) of the risks associated with the different levels of OPPT*PS¹

	Pregnant	Lactating	Empty	Pr.&La.
OPPT-1	0.463 (0.094)	0.061 (0.039)	0.591 (0.308)	- -
OPPT-2	0.132 (0.074)	0.038 (0.020)	0.326 (0.127)	0.018 (0.013)
OPPT-3	0.033 (0.017)	0.018 (0.009)	0.315 (0.148)	0.002 (0.001)

¹OPPT: Order of the positive pregnancy test; PS: Physiological state

For a given level (Pregnant, Lactating, Empty or Pregnant&Lactating) of physiological state animals with 1st order of positive pregnancy test (OPPT-1) always had the highest relative risk followed by does with 2nd (OPPT-2) and higher (OPPT-3) order of positive pregnancy test. Furthermore Table 3.2.4.2.1 shows that for given level (1st, 2nd or ≥ 3 order of positive pregnancy test) of OPPT the physiological state Empty was always the level with the highest risk followed by Pregnant, Lactating and finally Pregnant&Lactating. However because of the interaction between OPPT and PS the magnitude of the differences between levels of PS is different for different levels of OPPT and similarly the magnitude of the differences between levels of OPPT is different for different levels of PS.

In Figure 3.2.4.2.1 the posterior means of the risks associated with the different levels of the time dependent interaction between physiological state and number born alive, PS*BA, are plotted for each of the four levels of physiological state, PS. In general this figure shows the same pattern as Table III for PS, in the sense that for a given level of BA the highest risk is obtained for Empty followed by Pregnant, Lactating and Pregnant&Lactating. Only for level 2 of BA (1-2 young born alive) the physiological state with the highest risk is Empty followed by Lactating. This might be a sign of sick does. For these lactating does, after a poor kindling, the risk is dramatically increased due to the great physiological effort involved in lactation. For a given level of physiological state, the following is observed: For PS=Empty the highest risk is obtained for the third level of BA (3-4 young born alive) followed by the second level of BA (1-2 young born alive). For levels 4-8 of BA (5 or more young born alive) a tendency of decreasing risk with increasing number of young born alive was observed. Surprisingly the risk for level 1 of BA (0 born alive) was relative low. For PS=Pregnant the highest risk was observed for level 1 of BA (0 born alive). For levels of BA

different from 1, the risks were stable. For PS=Lactating the risk is decreasing with increasing number of young born alive. For PS=Pregnant&Lactating the risk associated with the different levels of BA are very low and stable. In rabbits, the pattern that risk is decreasing with increasing litter size was previously observed by Garreau et al. (2001) and Sánchez et al. (2004).

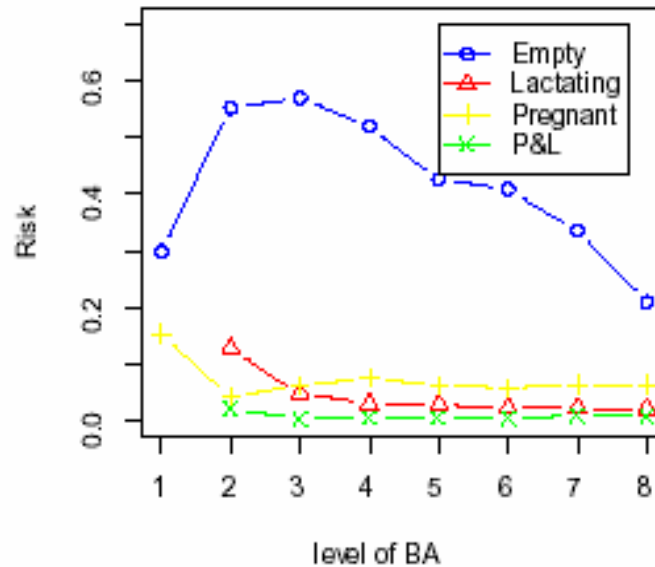


Figure 3.2.4.2.1.-Posterior means of risks associated with the different levels of PS*BA.

Table 3.2.4.2.2 shows the posterior standard deviations for the risk for all the levels of the interaction PS*BA.

Table 3.2.4.2.2.- Posterior standard deviations of the risks associated with the different levels of PS*BA¹

	Pregnant	Lactating	Empty	Pr.&La.
BA=0	0.095	--	0.143	--
BA=1-2	0.035	0.067	0.136	0.02
BA=3-4	0.037	0.027	0.319	0.005
BA=5-6	0.041	0.016	0.261	0.005
BA=7-8	0.033	0.015	0.209	0.004
BA=9-10	0.030	0.011	0.195	0.003
BA=11-12	0.033	0.011	0.164	0.006
BA= \geq 13	0.033	0.011	0.108	0.006

¹PS: Physiological state; BA: Number born alive.

3.2.4.3.-Baseline hazard function

Based on posterior means for each of the steps in the baseline hazard function the posterior mean of the cumulative baseline hazard function was computed. If the baseline hazard function is that of a Weibull distribution, then the natural logarithm of the cumulative baseline hazard function is a linear function of natural logarithm of time. This is not the case for longevity of our rabbit does, see Figure 3.2.4.3.1. Two separate regions can be observed, before and after the first parturition, around 19 days after the first positive pregnancy test ($\log(19) \approx 2.94$). The same conclusions were obtained by Sánchez et al. (2004), using the crude Kaplan-Meier estimator for the survival function, although in this previous work the beginning of the productive life was at the first mating.

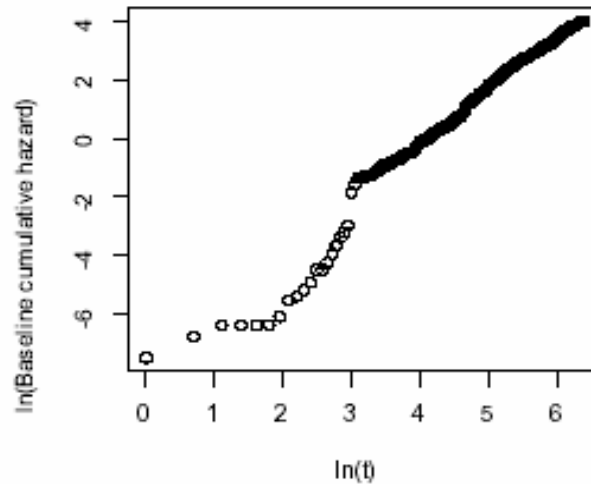


Figure 3.2.4.3.1.- Log posterior mean of $H_0(t)$ (Baseline cumulative hazard) versus $\ln(t)$ (days).

3.2.5.-Discussion

This study shows there is additive genetic variation for longevity of rabbit does, and thereby it may be possible to improve longevity by genetic selection. The marginal posterior mean of the additive genetic variance of rabbit does longevity was 0.247 and the 95% highest posterior density region was 0.046-0.482 (Table 3.2.4.1.1). However, despite the additive genetic variation, the marginal posterior mean of h_{\log}^2 (this heritability takes into account all the phenotypic variation on the $\log(g_i(t))$ scale) was 0.095, which is a low value. This value could be biased since in the analysis it was not

taken into account the fact that our population was being selected for litter size at weaning. However the genetic correlation between litter size at weaning and functional longevity has been estimated as very low and not significantly different from zero (Sánchez et al., 2005), thus, this bias is expected to be negligible. In rabbits, low values of heritability of longevity were also obtained by Garreau et al. (2001), but with a different definition of the trait and another survival model.

The marginal posterior mean of the variance of the residual effect in log-frailty was 0.690 which is almost 3 times higher than the additive genetic variance. This implies that the individual heterogeneity due to environmental unobserved effects and non additive genetic effects is more important compared with the individual heterogeneity due to additive genetic effects.

The data set in this analysis was previously analyzed using a different model, a semiparametric log-Normal animal frailty model without residual in log-frailty using the Survival Kit (Ducrocq and Sölkner, 1998), and with the covariates given in the present paper. With this model the marginal posterior mean of σ_a^2 was 0.204. This estimate is 17.4% lower than the estimate obtained from the animal model applied in this paper (0.247). These differences could partially be explained by a scale effect, since in the model implemented using the Survival Kit the total variance on $\log(g_i(t))$ scale was 1.849 ($Var(a + \varepsilon)$) and in the model with residual in log-frailty it was 2.582 ($Var(a + e + \varepsilon)$). Thus, almost the same figures for the h_{\log}^2 estimates (which do not depend on the scale) are observed. However, Damgaard (2004) observed in a simulation study that some bias in the estimation of model parameters exists when there is further individual heterogeneity, but it is ignored. More researches are needed in order to clarify to what extent the differences between the estimated values of model parameters are due to a scale effect or they are due to bias induced by do not consider further individual heterogeneity.

The estimates of the effects associated with the different levels of the covariates (β) were larger in absolute value in the animal model with residual term in log-frailty (marginal posterior means) than the ones from the animal model without residual term in log-frailty (joint posterior mode). The marginal posterior standard deviations for β were larger in the model with residual term in log-frailty than the standard error

reported by the Survival Kit. According to previous studies (Damgaard, 2004) these results could be expected and are in agreement with the scale effect.

To our knowledge there is only one study in which the animal model is implemented using the Survival Kit to infer the genetic variance of a longevity trait (Díaz et al., 2002). In order to check the behaviour of the Cox program of the Survival Kit in the estimation of additive genetic variance in animal models a simple simulation study was carried out. We simulated a population with the same structure as ours, using the model (1.1) but without residual in log-frailty and fixing the ordered survival times to every two days (to define the steps in the baseline hazard function). The model included a time-independent covariate with two levels and an additive genetic effect with a variance of 0.2. Thirty five replicates were generated, with a censoring level ranging from 45% to 55%, and the posterior modes, the posterior means and the posterior standard deviation of the genetic variance were estimated. The mean for the posterior means was 0.200 and their standard deviation was 0.050, the mean for the posterior modes was 0.181 and their standard deviation was 0.049, and the mean of the posterior standard deviations was 0.067 and their standard deviation was 0.009. Thus, the posterior mean could be considered an unbiased estimator of the additive genetic variance and the posterior mode as a downward biased estimator of the additive genetic variance in animal models. We have to mention that in the case of rabbit populations the females are normally closely related and this situation is not the case in other species where the females are not closely related (dairy cattle). Thus, it seems that this particularity of the rabbit populations makes possible to use the Laplacian integration implemented in the Survival Kit to infer the additive genetic variance using animal models.

One drawback with a Bayesian analysis using Gibbs sampling is often the computing time. The application in this study with 2,400 records of longevity, corresponding to around 50,000 elementary records and a pedigree file of 3,031 animals, took 10.8 days of CPU time in a SGI Altix 3,000 system. Thus, we have illustrated it is possible to carry out a Bayesian analysis using Gibbs sampling in the semiparametric log-Normal animal frailty model with time dependent covariates and a residual effect in log frailty. An advantage of using animal models for survival data is that all additive genetic relationships are taken into account, and biased estimation due to incorrect assumptions concerning the population structure is avoided. With still

increasing computer power, animal models for survival data can be expected to be applicable on large data sets within the near future.

3.2.6.-Acknowledgements

This work has been supported by the Spanish Research Project AGL2004-02710/GAN (CYCIT). Juan Pablo Sánchez thanks the Ministry of Education and Science of Spain for the grant to stay at the Department of Genetics and Biotechnology of the Danish Institute of Agricultural Sciences. The authors thank Mr. Christopher Jones for revising the English version of the manuscript.

3.2.7.-References

Adaptive rejection sampling (to generate random variables) from log-concave density functions. (42/1993) <http://lib.stat.cmu.edu/apstat/287> [consulted: 27 May 2005].

Andersen A.H., Korsgaard I.R., Jensen J., Identifiability of parameters in - and equivalence of animal and sire models for Gaussian and Threshold characters, traits following a Poisson mixed model, and Survival traits. Research Report 417 (2000). Department of Theoretical Statistics, University of Aarhus, Denmark.

Andersen P.K., Borgan Ø., Gill R.D., Keiding N., Statistical Models Based on Counting Processes, Springer, New York, 1992.

Diaz C., Chirinos Z., Moreno A., Carabaño M.J., Preliminary analysis of functional longevity in the Avileña Negra Ibérica beef cattle breed. In: Proc. of the 7th World Congress on Genetics Applied to Livestock Production, 19-23 August 2002, Vol. 29 Montpellier (France), pp. 697-700.

Damgaard L.H., Quantitative genetic analysis of survival, linear Gaussian and ordered categorical traits. Ph.D. thesis, 2004, The Royal Veterinary and Agricultural University, Frederiksberg C (Denmark).

Damgaard L.H., Korsgaard I.R., Simonsen J., Dalsgaard O., Andersen A.H., Genetic versus non-genetic Weibull log-normal sire frailty models: A simulation study.

In: Proceedings of the 2003 interbull technical workshop, Belstville, (USA).
Bulletin No 30 (2003), pp. 35-42.

Ducrocq V., Casella G., A Bayesian Analysis of mixed survival models. *Genet. Sel. Evol.* 28 (1996), 505-529.

Ducrocq V., Sölkner J., The Survival Kit V3.0, a package for large analyses of survival data. In: Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, 11-16 January 1998, Vol 27, University of New England, Armindale, pp. 447-450.

Estany J., Baselga M., Blasco A., Camacho J., Mixed model methodology for the estimation of genetic response to selection in litter size of rabbit, *Livest. Prod. Sci.* 21 (1989) 67-75.

Garreau H., Larzul C., Ducrocq V., Analyse de Longévité de la souche de lapins INRA 1077, in: Proceedings of the 9èmes Journ. Rech. Cunicole , 2001, Paris, France, pp. 217-220.

Geman S., Geman D., Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Analysis and Machine Intelligence* 6 (1984) 721-741.

Geweke, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. Oxford University Press, Oxford, 1992.

Gilks W.R., Wild P., Adaptive Rejection Sampling for Gibbs Sampling, *Appl. Stat.* 41 (1992) 337-348.

Korsgaard I.R., Madsen P., Jensen J., Bayesian Inference in the semiparametric log normal frailty model using Gibbs sampling, *Genet. Sel. Evol.* 30 (1998) 241-256.

Korsgaard I.R., Andersen A.H., Jensen J., 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 (2002) 307-333.

Korsgaard I.R., Andersen A.H., Sánchez J.P., Damgaard L.H., Identifiability of regression parameters for censored survival data. In preparation.

Ramón J., Rafel O., 1991-1992. Diez años de gestión global en España, in: *Proceeding of Expoaviga, X Jornada Cunícola, 2002, Barcelona, Spain, pp.113-117.*

Rosell J.M., Health status of commercial rabbitries in the Iberian peninsula. A practitioner's study, *World Rabbit Sci.* 11 (2003) 157-169.

Sánchez J.P., Baselga M., Peiró R., Silvestre M.A., Analysis of factors influencing longevity of rabbit does, *Livest. Prod. Sci.* 90 (2004) 227-234.

Sánchez J.P., Baselga M., Ducrocq V., Genetic and environmental correlations between longevity and litter size in rabbits, Submitted to *J. Anim. Breed. Genet.*

Yazdi M.H., Rydhmer L., Ringmar-Celerberg E., Lundeheim N., Genetic study of longevity in Swedish Landrace sows, *Livest. Prod. Sci.* 63 (2000) 255-264.

3.3.- Genetic and environmental correlations between longevity and litter size in rabbits⁷

3.3.1.- Abstract

An estimation of the genetic and environmental correlations between prolificacy (number of young born alive and number of young at weaning) and functional longevity in a population selected for litter size at weaning since 1992 was carried out. The method is only an approximation, but suggests that genetically longevity and litter size are not antagonistic objectives in breeding programmes, because the estimated genetic correlations are not significantly different from 0.

Key words: Correlation, Litter size, Longevity

3.3.2.- Introduction

Nowadays, culling rate in rabbit production is very high, around 120% yearly (Ramón and Rafel, 2002). Does are mainly culled because they die or become sick. About 50% of the does culled are in the course of the first three kindlings (Rosell, 2003). This leads to a population structure heavily relying on young does (Rosell et al., 1996) which show lower reproductive performances (Gómez, 1994) and are more difficult to manage. Despite the low price of females, replacement is a negative factor in production because there is a period between when a doe becomes sick and the final culling during which a cage is occupied by a low productive animal. Moreover, because most of the culled does are young, they have not been profitable yet at culling or death time. Another negative point of the high replacement rate is that when the replacement animals are bought outside, a high culling rate implies a higher proportion of animals coming into the farm periodically, increasing the risk of health problems. Thus, to genetically improve the doe's longevity and to reduce replacement rate could be an interesting strategy. To do this, first we have to know the genetic determinism of the trait and its relationship with other traits, particularly with traits that are included in the selection objective in breeding programmes, in order to be able to predict the consequences of selection on one trait over the others. Estimates of heritability of

⁷ Submitted to Journal of Animal Breeding and Genetics:
Sánchez J.P., Baselga M., Ducrocq V., 2005. Genetic and environmental correlations between longevity and litter size in rabbits.

longevity were previously been reported by Garreau et al. (2001) and Sánchez et al. (2004), but until now there has been no estimate of genetic correlations with other traits.

In rabbit breeding programmes, selection in maternal lines is mainly focused on litter size at birth or at weaning, and in paternal lines on feed conversion, indirectly selected through daily gain (Baselga and Blasco, 1989). Recently, other traits such as longevity have been considered in selection objectives (Sánchez et al., 2004, Garreau et al., 2001).

Estimates of the genetic and environmental correlations between growth traits and prolificacy are scarce. One of the more recent estimates is due to García and Baselga (2002b). They found that the estimated genetic, non-additive plus permanent environmental and environmental correlations between several growth traits and number at weaning did not differ significantly from zero. They also studied the correlated response on growth traits after selection on litter size and they did not find any significant correlated response. Garreau et al. (2000) found low, although slightly positive, estimates of correlations between litter size and growth traits. Their study was performed in a population selected for growth rate and they did not find any significant genetic trend on litter size. In the selection experiment of the line INRA1077 in France, which has been selected on litter size at weaning since the late nineteen-sixties, they found an undesired correlated response in weight at weaning, which has consequences on adult weight. As a consequence, this line is nowadays selected using a criterion which includes two traits, litter size at birth and weight at 63 days of life (Rochambeau, 1998).

In order to predict the consequences on longevity of selection on litter size or growth rate, it is desirable to know the genetic and environmental correlations between them. This study aimed at estimating the genetic and the environmental correlation between two prolificacy traits (number of young born alive and number of young alive at weaning) and longevity.

3.3.3.- Material and Methods.

3.3.3.1.- Animals

Data for this study were collected from the beginning of 1992 to the end of 2001 in a selection nucleus of rabbits, placed in Sant Carles de la Rápita (Tarragona), Spain. The animals belonged to the V strain (Estany et al., 1989). These animals were selected for litter size at weaning and the genetic evaluation was carried out using a family

index. In this nucleus, the management system was semi-intensive; the does were mated 11 days post-partum, the young were weaned at 35 days, and a pregnancy check through abdominal palpation was carried out 12 days post-mating. For a suitable genetic evaluation of animals in the nucleus, common culling criteria in commercial farms were not considered; i.e., does with a low level of production or not very large intervals between parities were not culled.

Longevity was defined as the time between the first positive pregnancy test and death or culling of the doe. The data set consisted of 2,400 longevity records (50% censored), 15,235 kindlings (2,320 does) and 14,604 weanings (2,219 does), and the pedigree file included 3,017 records.

3.3.3.2.-Statistical Methods

One way to study longevity, taking into account censored data, the presence of time dependent covariates and the non normal distribution of the trait, is to rely on the survival analysis methodology, which is based on the modelling of the hazard function. Litter size is analysed using linear mixed models. Currently, there is no obvious way to perform a multiple trait analysis combining longevity (studied using survival analysis techniques) and other linear traits when there are non zero residual correlations between traits (Ducrocq, 1999). Thus, approximations must be used (Ducrocq et al., 2003). Recently, Tarres et al. (2005) tested the approximated procedure via simulation and, in general, they obtained unbiased results.

The proposed approach aims at summarizing the data in such a way that a simple linear animal model can be used for each trait. This implies the calculation of a single “pseudo-record” $y_{t,i}^*$ for each animal i and each trait t . This “pseudo-record” combines all individual records pre-adjusted for fixed effects and for random effects other than the additive effect. The resulting “pseudo-records” with their weights ($\omega_{i,t}$) can then be analysed with the following animal model:

$$y_{t,i}^* = \mu_t + a_{t,i} + e_{t,i} \tag{2.1}$$

with variances $Var[a_t] = A\sigma_{a,t}^2$; $Var[e_t] = diag \left\{ \frac{1}{\omega_{i,t}} \right\} \sigma_{e,t}^2$

In order to estimate the genetic correlation between the “pseudo-records” for different traits, a random regression model equivalent to (2.1) can be used in order to avoid the presence of heterogeneous residual variances:

$$\omega_{t,i}^{1/2} \cdot y_{t,i}^* = \omega_{t,i}^{1/2} \cdot \mu_{t,i} + \omega_{t,i}^{1/2} \cdot a_{t,i} + \omega_{t,i}^{1/2} \cdot e_{t,i} \quad (2.2)$$

with variances: $Var[a_t] = A\sigma_{a,t}^2$; $Var[\omega_t^{1/2} \cdot e_t] = I\sigma_{e,t}^2$, and covariances between two traits t_1 and t_2 : $Cov[a_{t_1}, a_{t_2}] = A\sigma_{a_{t_1}, a_{t_2}}$; $Cov[(\omega_{t_1}^{1/2} \cdot e_{t_1}), (\omega_{t_2}^{1/2} \cdot e_{t_2})] = I\sigma_{e_{t_1}, e_{t_2}}$

In order to estimate the genetic and residual correlations between the traits, model (2.2) is solved keeping constant the variances estimated with the correct models and the original data. Tarres et al. (2005) showed that these constraints were necessary to obtain unbiased correlation estimates.

3.3.3.3.- Calculation of “pseudo-records” and their associated weight

For born alive (BA) and number at weaning (NW), a two-trait repeatability animal model was used in order to estimate variance components, and the fixed and permanent environment effects necessary to pre-correct the data. The models for both traits included a year-season effect (ys), a physiological state effect (ps) (nulliparous, primiparous suckling, primiparous weaned, multiparous suckling and multiparous weaned), a random non-additive plus permanent environmental effect (p) and a random additive genetic effect (a). The REML variance components estimation was carried out using the VCE software (Groeneveld, 1998) and the effects estimates were obtained using the PEST software (Groeneveld, 1990).

The “pseudo-records” and their weights for BA and NW were calculated as:

$$y_{BA,i}^* = \frac{\sum_{z=1}^{n_{BA,i}} [BA_{ikjz} - (\hat{p}_{BA,i} + y\hat{s}_{BA,k} + p\hat{s}_{BA,j})]}{n_{BA,i}} \quad \omega_{BA,i} = n_{BA,i}$$

$$y_{NW,i}^* = \frac{\sum_{z=1}^{n_{NW,i}} [NW_{ikjz} - (\hat{p}_{NW,i} + y\hat{s}_{NW,k} + p\hat{s}_{NW,j})]}{n_{NW,i}} \quad \omega_{NW,i} = n_{NW,i}$$

where $n_{t,i}$ ($t=BA$ or NW) is the number of kindlings or weanings of doe i and \hat{p} , \hat{y}_s , \hat{p}_s are estimates of the permanent environment and fixed effects in the model.

For longevity (L), a Cox animal model was used to estimate the additive genetic variance of the trait, the fixed effects and the additive effects. The fixed effects included in the model were: a year-season effect (ys) (time dependent, changing every 3 months) and an interaction between cycle order (co) (time dependent with 3 levels; first, second and third and more) and physiological status (ps) (time dependent with 4 levels; pregnant (with positive palpation), suckling, non-pregnant, and pregnant and suckling). To be consistent with the objective of this work, the effect of litter size is not included in the model in contrast with other studies (Sánchez et al., 2004, Garreau et al., 2001). These analyses were carried out using the Survival Kit 3.0 program (Ducrocq and Sölkner, 1998).

The results of the survival analysis of the longevity are expressed in terms of the hazard, which is a function of time and measures the daily probability for a doe of being culled given she was alive the day before.

For longevity, the “pseudo-records” and their associated weights were calculated extending to Cox frailty models the approach of Ducrocq (2001) and Ducrocq et al. (2003) for Weibull models. By using this procedure the associate weight to the “pseudo record” for one animal is equal to the part of the information matrix coming from the data. These values were:

$$y_{L,i}^* = \frac{\delta_i}{\omega_i} - 1 + \hat{a}_i \qquad \omega_{L,i} = e^{\hat{a}_i} \cdot \sum_{T_f=1}^{f_i} \left[d_f \cdot \frac{e^{x'_i(T_f) \cdot \hat{\beta}}}{\sum_{j \in R(T_f)} e^{x'_j(T_f) \cdot \hat{\beta} + \hat{a}_j}} \right]$$

where δ_i is the indicator variable which shows whether the record of animal i is censored ($\delta_i = 0$) or not ($\delta_i = 1$), \hat{a}_i is the EBV of the animal i , T_f are the distinct ordered failure times in the whole population, f_i is the time when the animal i dies or is censored, d_f is the number of does culled at time T_f (to account for ties), $R_{(T_f)}$ is the group at risk at time T_f and $x'_i(T_f) \cdot \hat{\beta}$ is the estimate of the vector of all fixed effects

affecting animal i at time T_f . The analogy with the approximation proposed by Ducrocq (2003) is the fact that the weight $\omega_{L,i}$ is the cumulative hazard of animal i (using Tsiatis' suggestion (1981) for the Cox model).

To estimate the correlation between the traits using the multiple trait model (2.2) a modified version of the AIREMLF90 program (Miształ, 2002) was used. The modifications of the program consisted in 1) imposing constraints such that the variances of the traits are fixed and are not estimated, 2) modelling the variance-covariance matrices as functions of correlations, directly leading to asymptotic standard errors of the correlation estimates.

3.3.4.- Results and Discussion

3.3.4.1.- Heritabilities and correlations for litter size traits

Table 3.3.4.1.1 shows the estimated heritabilities and the ratios between the variance of non-additive plus permanent environmental effects and the phenotypic variance.

Table 3.3.4.1.1.- Estimate and standard error of the heritability (h^2) and of the ratio between the variance of non-additive plus permanent environmental effects and the phenotypic variance (p^2) for the two litter size traits.

TRAIT ^a	h^2	p^2
BA	0.15 (0.015)	0.10 (0.010)
NW	0.08 (0.011)	0.07 (0.008)

^a BA: Born alive, NW: Number at weaning.

Using the complete bivariate linear model, the estimated genetic correlation between BA and NW was 0.885 (0.023), the estimated correlation between non-additive plus permanent environmental effects was 0.695 (0.037) and the estimated residual correlation was 0.604 (0.004). When the estimation of the genetic and environmental correlations were performed using the “pseudo-records” and their weights the genetic correlation was 0.923 (0.011) and the residual correlation was 0.663 (0.010).

Our estimated values for the heritabilities are higher than previous estimates in another subpopulation of the V line reported by García and Baselga (2002a). The estimated value for the ratios between the variance of non-additive plus permanent environmental effects and the phenotypic variance, and the estimated correlations

(genetic and between non-additive plus permanent environmental effects) are similar to those previously reported by García and Baselga (2002a).

3.3.4.2.-Heritability of longevity

The result from the Cox animal model gave an estimate of additive genetic variance of 0.181 (posterior mean) which corresponds to a heritability of 0.099 on the $\log(g(t_i))$ scale (Korsgaard et al., 2002), where $g(t_i)$ is the integrated function of the baseline hazard and the hazard due to time-dependent covariates. This value is larger than the one previously reported, for the same population, by Sánchez et al. (2004) (0.053). These two values should be compared with caution since here the trait definition is different, the physiological state is defined in a different way, the number of young born alive in each parturition is not included as a covariate and in our case we use an animal model and Sánchez et al. (2004) used a sire-maternal grandsire model. Other estimates of the heritability for longevity were reported by Garreau et al. (2001), who studied this parameter using different models, and the estimated value under the most realistic model was 0.05. This figure is defined on a different scale but is still a low value.

3.3.4.3.-Validation of the “pseudo-records”

The validity of the “pseudo-records” for longevity was checked by computing the correlation between the EBVs from the Cox animal model and those based on the BLUP animal model (2.1) using the “pseudo-records”, their weight, the additive variance estimated with the Cox animal model and a residual variance of 1.0. The derivation of the “pseudo-record” formula was done aiming at a correlation close to 1.0. In our case, this correlation was 0.99934.

For BA and NW, the correlation between EBVs from univariate analysis, with the correct models and data, and the estimated ones using model (2.1) with the “pseudo-records”, their weight and the variance components estimated in univariate REML analysis, was 0.99998.

3.3.4.4.-Estimation of the correlations between litter size traits and longevity

Table 3.3.4.4.1 shows the variance components used in the analyses to estimate the correlations between BA and L, and NW and L, in two separate bivariate analyses.

These variance components were kept constant in the analysis of the multi-trait model (2.2). These values were estimated by single trait analyses for L (Survival Analysis) and by a bivariate analysis for BA and NW (REML).

Table 3.3.4.4.1.- Variance components used to estimate the correlations for the litter size traits and longevity

	BA	NW	L
σ_e^2	8.937	5.524	1.0
σ_a^2	1.766	0.546	0.181

σ_e^2 : Residual variance, σ_a^2 : Additive genetic variance, BA: Born alive, NW: Number at weaning, L: Longevity.

The genetic and residual correlations and their associated standard errors are shown in Table 3.3.4.4.2.

Table 3.3.4.4.2.- Genetic and environmental correlations between longevity and litter size traits (asymptotic standard error assuming that the variances used are correct).

	Genetic	Environmental
BA-L	0.16 (0.093)	-0.11 (0.016)
NW-L	-0.17 (0.114)	-0.01 (0.019)

BA: Born alive, NW: Number at weaning, L: Longevity.

Negatives values for the correlations represent a favourable relationship between longevity and the litter size trait, because increasing the litter size leads to a lower hazard and, thus, a higher longevity. Given that the standard error estimates for the genetic correlation were obtained assuming that the fixed variances were correct, they can be regarded as underestimates of the true asymptotic standard errors (Tarres et al., 2005). Consequently, the genetic correlations can be considered as not significantly different from zero. Regarding the environmental correlations only the parameter between BA and L can be considered as significantly different from zero being the estimate value favourable but very low (-0.11 (0.016)).

Other estimates of the correlation between longevity and litter size traits in rabbits were not found in the literature. Nevertheless, in two studies from selection experiments on litter size, the correlated response in longevity has been studied and may give an indication of the existence of genetic correlation between both types of traits.

Rinaldo and Bolet (1988) showed that after seven generations of selection on litter size at weaning, neither the number of litter per doe nor the number of days of productive life were affected. Theilgaard et al. (2005) showed that, in crossbred does, selection on litter size at weaning during 12 generations did not affect longevity. Our results are in agreement with these correlated responses to selection previously reported. They show that longevity and litter size are not antagonistic traits in rabbit breeding programmes. The selection on one of them is not expected to modify the other.

3.3.5.-Acknowledgements

This work has been supported by the Spanish Research Project AGL2004-02710/GAN (CYCIT). The authors thank Mr. Neil Macowan for revising the English version of the manuscript.

3.3.6.-References

- Baselga, M., Blasco, A. (1989). *Mejora genética del conejo de producción de carne*. Edit. Agrogúas mundi-prensa. Madrid, Spain.
- Ducrocq, V. (1999). Topics that may deserve further attention in survival analysis applied to dairy cattle breeding – some suggestions. In: *Proceeding of International Workshop on Genetics Improvement of Functional Traits in Cattle (GIFT); Longevity*. Jouy-en-Josas, (France). *Interbull Bulletin No 21*, pp. 181 – 189.
- Ducrocq V. (2001). A two-step procedure to get animal model solutions in Weibull survival models used for genetic evaluations on length of life. In: *Proceedings of the 2001 Interbull meeting*. Budapest, (Hungary). *Interbull Bulletin No 27*, pp 147-152.
- Ducrocq, V., Delaunay, I., Boichard, D., Mattalia S. (2003). A general approach for international genetic evaluations robust to inconsistencies of genetic trends in national evaluations. In: *Proceedings of the 2003 Interbull meeting*. Beltsville, (USA). *Bulletin No 30*, pp. 101-111.
- Ducrocq, V., Sölkner, J. (1998). The survival kit V3.0, a package for large analyses of survival data. In: *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production*. Armindale (Australia). Vol. **27**, pp. 447- 450.

- Estany, J., Baselga, M., Blasco, A., Camacho, J. (1989). Mixed model methodology for the estimation of genetic response to selection in litter size of rabbits. *Livest. Prod. Sci.* **21**: 7-75.
- García, M.L., Baselga, M. (2002a). Estimation of genetic response to selection in litter size of rabbit using a cryopreserved control population. *Livest. Prod. Sci.* **74**: 45-53.
- García, M.L., Baselga, M. (2002b). Estimation of correlated response on growth traits to selection in litter size of rabbits using a cryopreserved control population and genetic trends. *Livest. Prod. Sci.* **78**: 91-98.
- Garreau, H., Larzul, C., Ducrocq, V. (2001). Analyse de longévité de la souche de lapins INRA 1077. In: Proceedings of the 9^{èmes} Journ. Rech. Cunicole, Paris, (France). pp. 217 - 220.
- Garreau, H., Szendro, Zs., Larzul, C., De Rochambeau, H. (2000). Genetic parameters and genetic trends of growth and litter size traits in the White Pannon breed. In: Proceedings of 7th World Rabbit Congress, Valencia, (Spain). Vol. **A**, pp. 403 – 408.
- Gómez E.A. (1994). La selección del tamaño de camada en conejo de carne : Influencia de los efectos maternos y de la heterogeneidad genética entre partos. *Tesis doctoral. Universidad Politécnica de Valencia, Spain.*
- Groeneveld, E. (1990). Pest user's manual. Institute of Animal Husbandry and Animal Behavior, FAL. Germany.
- Groeneveld, E. (1998). VCE4 user's guide and reference manual, version 1.3. Institute of Animal Husbandry and Animal Behavior, FAL. Germany.
- 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.

Misztal, I. (2002). REMLf90 Manual.

<http://nce.ads.uga.edu/~ignacy/numpub/blupf90/docs/remlf90.pdf>

Consulted

29/08/2005.

Ramón, J., Rafel, O. (2002). 1991-2000. Diez años de gestión global en España. In: Proceedings of Expoaviga 2002, X jornada cunícola, Barcelona, (Spain). pp.113–117.

Rinaldo, D. and Bolet, G. (1988). Effect of selection for litter size at weaning on reproductive life of female rabbits. In: Proceedings of the 4th World rabbit congress, Budapest, (Hungary). Vol. 1, pp. 269-275.

de Rochambeau, H. (1998). La femelle parentale issue des souches expérimentales de l'INRA évolutions génétiques et perspectives. In: Proceedings of the 7^{èmes} Journ. Rech. Cunicole, Lyon, (France). pp. 3-14.

Rosell, J.M. (2003). Health status of commercial rabbitries in the Iberian peninsula. A practitioners study. *World Rabbit Sci.* **11**:157-169.

Rosell J.M., Badiola J.I., De la Fuente L.F. (1996). Rhinitis of the domestic rabbit. An epidemiological survey during the period 1992-1993. Influence of the sex, the age and the phase of lactation. In: Proceedings of the 6th World Rabbit Congress, Toulouse, (France). Vol 3, pp. 113-117.

Sánchez, J.P., Baselga, M., Peiró, R, Silvestre, M.A. (2004). Analysis of factors influencing the longevity of rabbit does. *Livest. Prod. Sci.* **90**: 227:234.

Tarres J., Piedrafita J., Ducrocq V. (2005). Validation of an approximate approach to compute genetic correlations between longevity and linear traits. *Genet. Sel. Evol.* (*in press*)

Theilgaard, P., Sánchez, J.P., Pascual, J.J., Friggens, N.C., Baselga, M. (2005). Effects of selection for prolificacy and body fatness on survival of rabbit does assessed using a cryopreserved control population. *Submitted to Livest. Prod. Sci.*

Tsiatis A. (1981). A large sample study of Cox's regression model. *Ann. Statist.*, **9**:93-108.

CHAPTER 4: CONSTITUTION AND EVALUATION OF THE L-P LINE

CHAPTER 4: CONSTITUTION AND EVALUATION OF THE L-P LINE

4.1.- Constitution of the L-P line

The constitution of the L-P (Long lived – Productive) was conducted selecting does at commercial farms excelling because of their high longevity but being above the mean in prolificacy. It was intended to apply a very high intensity of selection for longevity, in a similar way as it was done in rabbits (Cifre et al., 1998a,b) or pigs (Bichard and David, 1985; Sorensen and Vernersen, 1991; Herment et al., 1994; Noguera et al., 1997) in experiments and programmes of hyper-prolificacy.

Firstly, we had to define the so called L-P does. These females were animals which showed an extremely long productive life and kept a prolificacy near or above the mean. Approximately, between two and five does were selected per thousand of females screened at the commercial farms. The actual intensity of selection is not known because the longevity does not follow a normal distribution and, moreover, the selected does had censored data for longevity (they were still alive when they were selected). However, an idea of the effort of selection can be got through the minimum of longevity and prolificacy among the selected L-P does. These values were 25 parities and 7.5 young born alive.

The foundation of the L-P line took place in three steps (Figure 4.1.1). The first step, that tried to get sons of L-P does, was carried out during April 2002. A total of 15 L-P does were detected in 8 commercial farms. These animals were placed in a small lazaretto close to the facilities of the Animal Science Department of the Polytechnic University of Valencia (UPV). The does were inseminated with semen from bucks of the current generation (27) of the V line. Only 12 of the 15 recruited females had male offspring, in total 15 males. The mean of the number of parities of these 12 does was 30.8, and their prolificacy mean was 8.7 young born alive.

The second step tried to get sons of a new batch of LP does mated to the males got in the first step. Since September 2002 until November 2002 another set of 15 L-P does was selected and housed in the same lazaretto than the previous batch. These does were inseminated with semen coming from the bucks obtained in the first step and a total of 17 males were got. These males were grandsons of 11 of the L-P does of the first batch. Thus, one L-P doe having sons did not have any grandson. The mean of the

number of parities of the 11 L-P does, dams of the second crop of males, was 33.9, and their prolificacy mean was 8.8 young born alive.

The third step in the constitution of the L-P line was started in June 2003 and, until November 2003, 32 L-P does were detected in 25 farms of Spain and Portugal. Again they were housed in the facilities of the UPV and were inseminated with semen coming from the 17 bucks obtained in the previous step. Offspring (males and females) was obtained from 30 of these 32 does, and this offspring was the initial generation of the L-P line. The mean number of parturitions of the 30 L-P does which left offspring was 28.8, and the mean number of young born alive was 9.0. This zero generation was constituted by 42 does and 32 bucks. The bucks represented 27 of the final 30 L-P does which left offspring and the other three origins up to 30 were recovered by female way.

In all the steps of the foundation, the L-P recruited does were inseminated and they were slaughtered on the 29th day of gestation and the non-born young were disinfected and fostered to does that had the parturition during the same week, at the main facilities of the UPV. This was done in order to avoid the contact of the young with the L-P does because afterwards they would be transferred to the main facilities of the UPV and, thus, sanitary problems would be avoided.

Once the zero generation of the L-P line was founded, it was necessary to evaluate the response to selection. In order to achieve this objective we designed an experiment in which the performances for longevity and reproductive traits were compared between the new L-P line and the V line (Estany et al., 1989). V line is a maternal strain which has been selected during 31 generations for litter size at weaning. The zero generation of the L-P line was random mated (with the restriction of not mating full or half sibs) to form an experimental population of approximately 250 females. Another 250 females from V line were randomly chosen from the generation 31. The V line was chosen as control because it was difficult to obtain a non selected control population from the origin population of the L-P females and, because the V line is recognised as a very well performing maternal line, it represents a high standard for comparison.

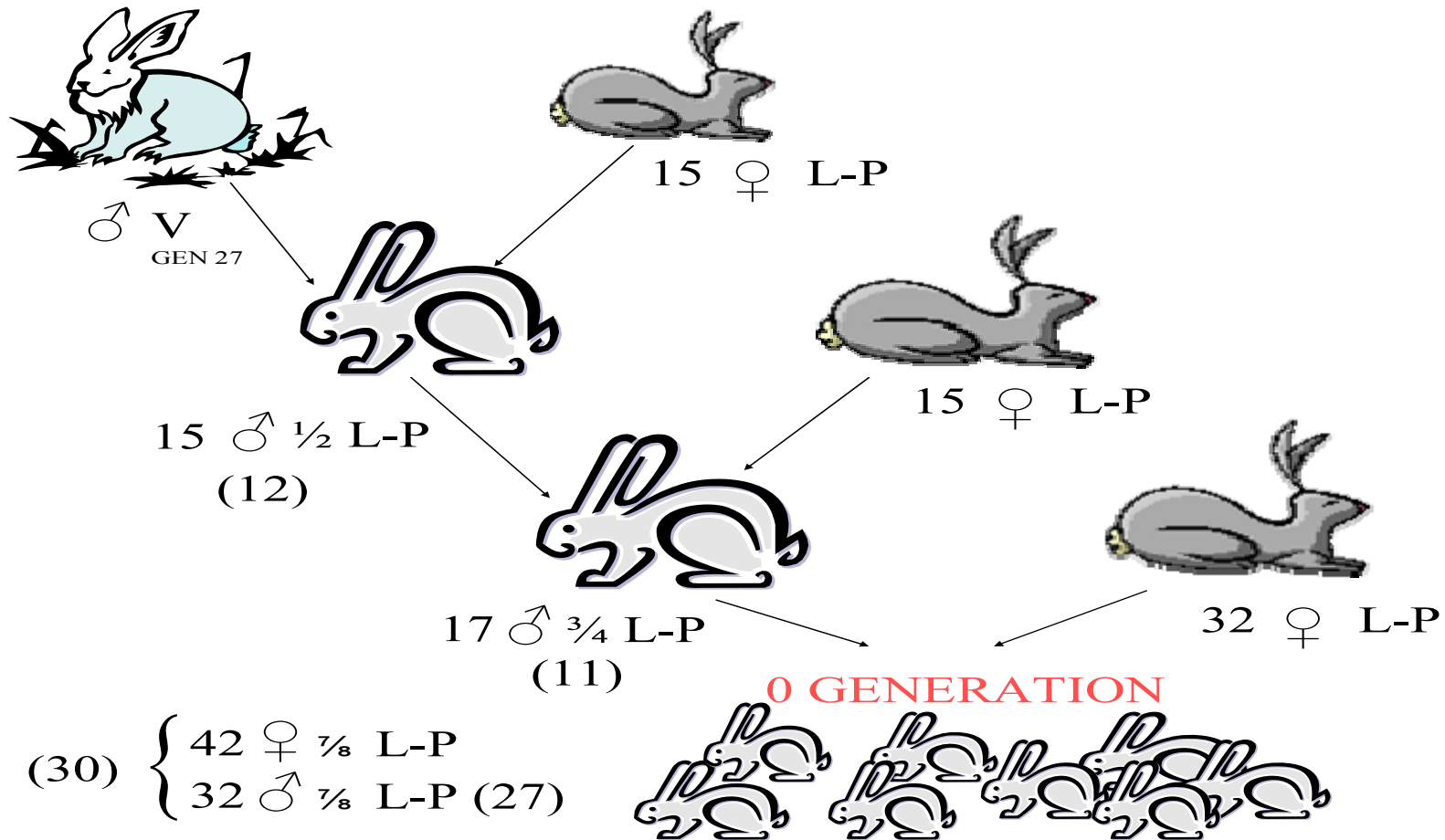


Figure 4.1.1.- Constitution of the L-P (Long lived - Productive) line. In each step appears the number recruited of L-P does and the total number of male offspring obtained (both sexes in generation 0). In brackets the number of L-P females, actual dams of the offspring indicated.

4.2.- Experimental design

The experimental design was carried out focusing on the longevity trait, which was the main trait in the selection procedure. In Collett (2003) a detailed description can be found on how to determinate the size of an experiment aiming to detect differences in survival in two distinct populations.

Assuming a proportional hazards model for the survival times, the hazard of death at time t for a L-P animal will be proportional to the hazard of one animal in the control line (V line):

$$h_{L-P}(t) = \gamma_R h_V(t)$$

Where, $h_{L-P}(t)$ is the hazard of the L-P population, $h_V(t)$ is the hazard of the control population and γ_R is the unknown hazard ratio between this two populations. θ_R could be defined as $\log(\gamma_R)$, so if θ_R is equal to 0 this means that there are not differences between the two populations, negative values for θ_R indicate that the survival is higher for the L-P population and positive values indicate that the animals in the V line live longer.

Thus, the objective is to determine the number of animals needed in the experiment to be able to test the null hypothesis $H_0 \equiv \theta_R = 0$ with the type I (α) and type II (β) errors desired, using the log-rank test (see section [1.2.3.2.2](#)). For this objective, firstly, the number of deaths needed has to be computed and, finally, knowing the survival functions of the populations to determine the number of animals to include in the experiment.

The required number of deaths is taken to be such that there is a probability of $1 - \beta$ (the power of the test) of declaring the real log-hazard ratio θ_R to be significantly different from zero.

The required number of deaths, d , can be obtained from the equation

$$d = \frac{4 \cdot c(\alpha, \beta)}{\theta_R^2} = \frac{4 \cdot (Z_{\alpha/2} + Z_{\beta})^2}{\theta_R^2}$$

Where, $Z_{\alpha/2}$ and Z_{β} are the ordinates of the standard Normal distribution at points $\alpha/2$ and β . Thus the computation of d requires the definition of: α , β , and the expected value for the real log-hazard ratio (θ_R).

The derivation of d can be found in detail in Collett (2003). It is based on the fact that the log-rank test is used to determine the significance of the hazard-ratio between the two populations, and on the fact that the log-rank statistic has an approximate Normal distribution.

Once the number of deaths required is known, the number of individuals to start the trial has to be computed. To do this we have to determine the probability of death over the duration of the trial. Thus, n (the number of animals needed in the study) is:

$$n = \frac{d}{P(\text{death})}$$

The probability of death is computed as:

$$P(\text{death}) = 1 - \frac{1}{6} \{ \bar{S}(f) + 4\bar{S}(0.5a + f) + \bar{S}(a + f) \}$$

Where, a is the duration of the accrual period (the period during which the animals are included in the experiment), f is the duration of follow-up period (the duration of the experiment) and $\bar{S}(t)$ is the mean survival function of the two populations involved in the study. The derivation of the probability of death can also be found in Collett (2003).

For our experiment the accrual period was fixed to 90 days and the follow-up period to 365 days. Moreover, the survival function of the V line was the one estimated in the section [3.1.4.2](#). Several expected log-hazard ratios, type I and type II errors were studied. The survival functions for the L-P line were computed based on the one from the V line taking into account the different expected log-hazard ratios, and assuming proportionality. Figure 4.2.1 shows the expected survival functions for the two log-hazard ratios studied ($\theta_{R1} = -0.31$ and $\theta_{R2} = -0.15$). These log-hazard ratios are equivalent to the hazard ratios $\gamma_{R1} = 0.73$ and $\gamma_{R2} = 0.86$. These differences imply an improvement in the mean survival of 48 days (13.2%) on one hand ($\theta_{R2} = -0.314$) and 24 days (6.5%) on the other ($\theta_{R2} = -0.15$). If the mean is estimated restricted to 330 days (latter we will use this estimation of the mean) the improvement is 19 days (7%) on one hand ($\theta_{R2} = -0.31$) and 9 days (4%) on the other ($\theta_{R2} = -0.15$).

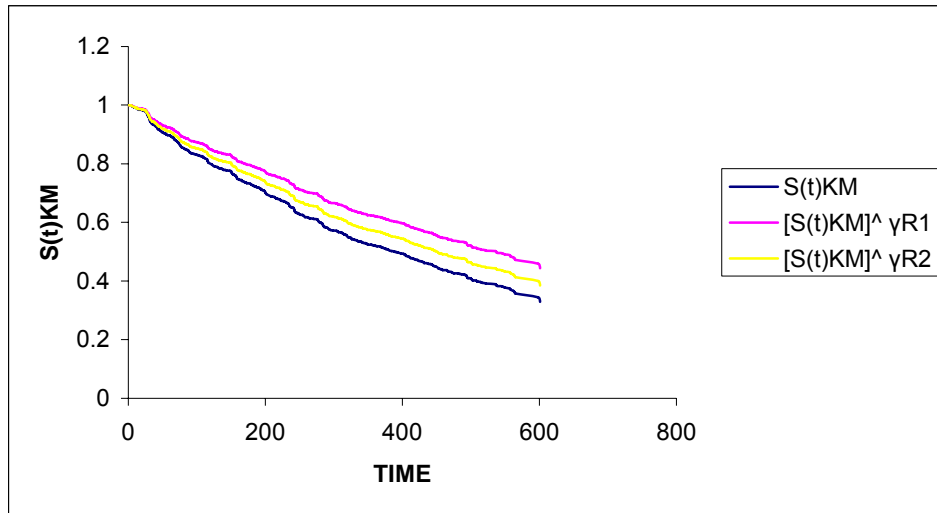


Figure 4.2.1.- Survival functions for the reference population (line V) ($S(t)_{KM}$), and the two populations with different hazard ratios ($[S(t)_{KM}]^{\gamma_{R1}}$) and ($[S(t)_{KM}]^{\gamma_{R2}}$).

Table 4.2.1 shows the number of animals needed to detect as significantly different from zero, different expected log-hazard ratios (θ_R) (which implies different death probabilities), for different type I and type II errors.

Table 4.2.1.- Number of animals required to detect as significant two different log-hazard ratios for different Type I and Type II errors

θ_R	P(death)	Type I error	Type II error	N° of animals
-0.31	0.469	0.05	0.1	908
			0.2	680
			0.3	533
		0.1	0.1	739
			0.2	534
			0.3	407
-0.15	0.500	0.05	0.1	3736
			0.2	2791
			0.3	2193
		0.1	0.1	3043
			0.2	2197
			0.3	1674

θ_R = expected log-hazard ratio

These theoretical results were assessed by means of simulation. We were interested in studying if the theoretical equations gave biased estimates of the power of the experiment (or the type II error). The twelve cases of table 4.2.1 were simulated and in each one, the distribution of animals between the two populations was balanced. The

survival times in the twelve cases were simulated from a Weibull distribution with ρ equal to 1.04 and $\rho \log(\lambda)$ equal to -6.59. This combination of Weibull parameters was chosen to have a median survival of 386 days. Survival times larger than 365 days were censored at this time and the censoring level ranged from 55 % to 65 %. In each population this design was replicated 100 times and in each one the long-rank statistic and its corresponding p-value were computed. Afterwards, the times in which the null hypothesis (log-hazard ratio equal to 0) was rejected, with the desired type I error, were counted. The results of the simulation are showed in table 4.2.2.

Table 4.2.2.- Simulation check of the statistical power of the different experimental sizes for the two log-hazard ratios assumed

θ_R	N	N° p-values ^a ≥ 0.05	N° p-values ≥ 0.1
-0.13	908	14	--
	680	27	--
	533	30	--
	739	--	19
	534	--	22
	407	--	33
-0.15	3736	14	--
	2791	23	--
	2193	29	--
	3043	--	11
	2197	--	22
	1674	--	35

θ_R = expected log-hazard ratio, ^a Absolute values but because the number of replicates were 100 also can be interpreted directly as percentages.

In table 4.2.2, the times in which the p-values are greater than the reference value for the type I error (0.05 or 0.1), represent the times in which the null hypothesis is not rejected when in fact is false (type II error). The results from the simulation show that the previous values reported in the Table 4.2.1, calculated using the theoretical equations, slightly underestimate the type II errors.

Due to the capacity of our facilities the size of the experiment was fixed to approximately 500 females, this size corresponds to a power of between 0.7 ($\alpha=0.05$) and 0.8 ($\alpha=0.1$). However, the design was defined assuming that the two populations were allocated in the same installations, although the real experiment was developed in three farms. This fact could reduce the power of the experiment and, moreover, the animals are actually related and this also will reduce the power of the experiment. After

these considerations it seems sensible to fix $\alpha=0.10$ for the test of difference of longevity in our experiment.

4.3.- Evaluation of the L-P line

4.3.1.- Abstract

An evaluation of a new maternal line (L-P) of rabbits was carried out. This new line was founded following a schema similar to that applied in the selection for hyper-prolificacy in rabbits or pigs. In this case the selection criteria were hyper-longevity and independent culling levels near the average for prolificacy. The evaluation was carried out by the comparison of reproductive and longevity performances of the L-P line to other maternal line with recognized good reproductive performances and standard longevity (V). At early ages the survival ability of the L-P line did not seem to have been improved, however at late ages an improvement appeared. Regarding reproductive traits, the V line showed a better performance than the L-P line, especially during the first parturitions, but later in life these differences disappeared. If the comparison is done averaging the differences between all the parturitions, the V line showed around 0.5 young more than the L-P line. In spite of the better prolificacy of the V line it did not show a higher cumulative production than the L-P line since the latter showed a slightly better longevity and there was no difference in fertility between them.

Keywords: rabbit breeding, lines foundation, longevity, prolificacy, fertility, cumulative production.

4.3.2.- Introduction

The annual replacement rate of females in rabbits for meat production is very high (120 %, Ramón and Rafel, 2002) and to study how to increase the length of productive life of the females could be a very interesting research area. Previous studies have shown that the heritability of this trait is between 0.05 (Garreau et al., 2001; effective heritability) and 0.1 (Sánchez et al., 2005; on log-g(t) scale), thus selective breeding could be an alternative to achieve this objective.

Nowadays, standard rabbit breeding programs are based on the selection within lines for the current characters of interest (prolificacy or growth). Regarding the prolificacy traits, the size of the lines under selection ranges between 100-250 does and 20-40 bucks, and the generational interval is around 9 months (Baselga, 2004). These short generational intervals are achieved because the information used to evaluate the candidates comes from their own performance in the first parturitions (females) and from collaterals and ancestors but there is no phenotypic information from offspring.

Under this structure of the information it does not seem feasible to include the longevity as another selection objective, because the accuracy of the evaluation would be very low. In order to improve the accuracy, for a given population size, the generational interval should be increased and also the usage of reproductive techniques (cryoconservation of gametes and embryo transfer) would be necessary.

An alternative to improve the accuracy of the genetic evaluation without increasing the generational interval is to include information from other traits, early recorded and genetically correlated with longevity. In spite of the theoretical difficulties for a multi-trait approach, involving one survival trait and linear Gaussian or threshold characters, having been theoretically solved by Damgaard (2004), the practical implementation of these models has the important limitation of enormous computation times. Thus, approximations have been proposed (Ducrocq, 2003; Tarres et al., 2005) to estimate correlations between longevity and other traits and to include information from other correlated traits in the genetic evaluation of longevity. Unfortunately in rabbit breeding, nowadays there is a complete lack of knowledge about the correlations between early recorded traits and longevity. Perhaps the main reason is that the methodological procedures to estimate these correlations have been proposed very recently.

The heritability for longevity can be considered low, as for prolificacy. The within line response to selection for prolificacy has been estimated and ranges from 0.076 (Tudela et al., 2003) to 0.085 (García and Baselga, 2002) young per parturition and generation, and these figures are considered lower than expected. Thus, in the hypothetical case of selection within line for longevity lower responses could be expected.

Alternative approaches have been used to improve prolificacy traits. One of them was inspired by the successful results of the programs and experiments of hyper-prolificacy in pigs (Bichard and David, 1985; Sorensen and Vernerisen, 1991; Herment et al., 1994; Noguera et al., 1997). Thus, applying very high intensities of selection for number born alive, in commercial populations of rabbits, a competitive new maternal line of rabbits was founded (Cifre et al., 1998 a, b). A similar approach, trying to improve the longevity, has been carried out by us and a new line has been founded, searching in commercial rabbit farms outstanding does for their longevity that had prolificacy near or above the average.

The objective of this study is to check whether this approach has been successful, comparing the new line to a well reputed maternal line of high prolificacy and standard longevity.

4.3.3.- Material and Methods

4.3.3.1. - Animal Material and Management

The two lines compared in this experiment were the V line and the L-P line, the former has been selected for litter size at weaning for 31 non overlapping generations. This line was founded in the early 1980s (Estany et al., 1989) from different synthetic populations. The L-P line has been recently founded selecting females in commercial farms (L-P does) which showed an extremely high productive life (measured as number of parturitions) but keeping a prolificacy equal or above the population average during their whole life (measured as the mean number of young born alive per parturition). The foundation took place in three steps. Table 4.3.3.1.1 shows the summary statistics of the selected does in each step for the traits of interest. In the first step the selected does were mated to V bucks, in the second step the male offspring from the previous step were mated to a second batch of L-P females and in the third step the male offspring from the second step were mated to a third batch of L-P does. The offspring of this step were randomly mated to produce the experimental population.

Table 4.3.3.1.1.- Summary statistics for the number of parturitions and for the mean number of young born alive of the females which contributed to the foundation of the L-P line.

	Females with offspring	Parturitions			Means of born alive		
		Mean	Max.	Min.	Mean	Max.	Min.
Step 1	12	30.8	38	25	8.7	10.3	7.5
Step 2	11	33.9	41	28	8.8	9.7	7.8
Step 3	30	28.8	36	25	9.0	11.9	7.9

A total of 570 females were raised for the experiment (282 V (generation 31) and 288 L-P). They were housed in three different farms (I (76 V and 73 L-P), II (77 V and 73 L-P) and III (129 V and 142 L-P), trying to balance the representation of the different groups of full sibs in the farms. Some of these does died before the initiation of its productive life (first positive pregnancy test) and they were not included in the analysis. The experiment took place from October 2004 to November 2005.

In farms I and II the mating was by artificial insemination with semen from other paternal line but in farm III the mating was natural, using bucks from the respective lines. The first mating took place when the does were 4.5-5.5 months old and the females were again mated 25 days after parturition. In all the farms the pregnancy test was carried out by abdominal palpation 12 days after mating. In farms I and II the non pregnant does were re-mated 28 days after the previous insemination. In farm III the does which did not accept the buck were presented to the male one week later and the does which accepted the male but did not become pregnant were re-mated 21 days after the previous presentation. In the three farms a control of the number of young alive 28 days after parturition was conducted and this was considered as the number of young at weaning. However, in farms I and II the real weaning took place between 35 and 40 days after parturition.

In the three farms the same culling policies were implemented. Does with low prolificacy were never culled. In addition to the common practice of culling due to evident pathological problems (i.e.: Snuffles, Mastitis, Sore hocks, Diarrhoea, etc), does with three consecutive non fertile matings (in all the farms), or with six consecutive refusals to the buck (in farm III) were culled. Also, the does that after two consecutive pregnancies did not have any young alive at weaning were also culled. All these reasons for culling are considered indicators of abnormal reproductive troubles but not indicators of poor production of healthy animals.

4.3.3.2. – Traits Definition and Statistical Models

4.3.3.2.1.- Functional Longevity

The longevity was defined as the time in days from the first positive pregnancy test until the death or culling of the female. All the females alive at the end of the experiment had a right censored record for longevity. This trait could be considered as a measurement of the functional longevity since no culling based on production is carried out.

This trait was analysed using a semi-parametric log-Normal animal frailty model in a Bayesian MCMC implementation (Sánchez et al., 2005). The factors included in the model were:

i.- A time-dependent combination between the number of young born alive in each parturition and the physiological status. The number of young born alive was recoded in 9 levels (1st, zero born alive; 2nd, 1-2 born alive; 3rd, 3-4 born alive, ..., 8th,

≥ 13 born alive and 9th include all the does before the first parturition). The changes of level for this effect took place at the parturition date. The physiological status had 4 levels (1st, pregnant; 2nd, lactating; 3rd, unknown and 4th, empty). A doe was considered to be pregnant after a positive pregnancy test, lactating after the parturition if it had at least one born alive, unknown when the doe was mated but was not already tested for pregnancy, empty after parturition with zero born alive or after a negative pregnancy test).

ii.- A time-dependent combination between the physiological status and the number of cycle (5 levels: 1st, between the first and the second positive pregnancy test; 2nd, between the second and the third positive pregnancy test, and so on until the 5th which include does with 5 or more positive pregnancy tests).

iii.- A time-dependent combination between the genetic type (2 levels: L-P and V lines) and the number of cycle .

iv.- A time-dependent combination between the farm (3 levels: I, II and III) and the year-season effect (6 levels: the first include a period of three months and the rest two-months periods).

v.- A time-independent log-frailty additive genetic effect, which was assumed to be normally distributed with mean 0 and variance $A\sigma_a^2$, being A the numerator relationship matrix between the individuals.

vi.- A time-independent log-frailty residual effect, which was assumed to be normally distributed with mean 0 and variance $I\sigma_e^2$.

An improper uniform prior was assumed for the covariate effects. Multivariate Normal priors were assumed for the additive genetic and residual effects, and they were assumed to be independent of each other, and of the covariate effects. A mass of probability was assigned at points 0.247 and 0.69 for the residual and additive genetic variances. These figures were those previously estimated by Sánchez et al. (2005) for the V line.

After a burn in period of 200,000 rounds and retaining one of each 200 sampled values the convergence of the Markov chain (total length 2,000,000) was assessed by means of the Geweke test for the contrasts of interest.

In a previous study to design our experiment and to check its power, we observed that given a population of our experimental size of non related animals and the genetic type as the only factor, we could detect as significant an improvement of 48

days (13.2 %) in the mean life with a power of 0.8 and a type I error of 0.1. Thus, it seems sensible to fix the type I error to 0.1 for testing larger differences in longevity between both lines, since actually the animals are relatives and more than one classification factor is included into the model, reducing the real power of the experiment. Because of this, under a Bayesian approach, we will report 90% highest posterior density regions (HPD) instead of 95% HPD.

The results from the survival analysis can be expressed in time units to better understand what the log-hazard or hazard ratios imply. This can be achieved by using the restricted (to the higher uncensored observation) mean life estimate. Thus, the survival functions up to 330 days (the highest uncensored observation) were computed for five pairs of animals; within each pair one belonging to the V line and the other to the L-P line. For the first pair, the survival functions ($S_{j,1}(t)$, for $j = \text{L-P, V}$) were predicted using a combination of covariates with a very low associated hazard (they lived in a favourable farm-year-season, never entered in the empty physiological state and in all the parturitions had 10 young born alive). For the second pair, the survival functions ($S_{j,2}(t)$, for $j = \text{L-P, V}$) were predicted using a combination of covariates with a medium associated hazard (they lived in a favourable farm-year-season, needed two matings for all their positive pregnancies and in all the parturitions had 10 young born alive). For the third pair, the survival functions ($S_{j,3}(t)$, for $j = \text{L-P, V}$) were predicted using a combination of covariates with a high associated hazard (they lived in an unfavourable farm-year-season, alternated between cycles in which they became pregnant first time and cycles in which they needed three matings for a positive pregnancy, and in all the parturitions had 10 young born alive). Based on the predicted survival functions for the first pair, the survival functions for another two pairs were predicted as:

$$S_{j,i}(t) = S_{j,1}(t)^{\exp\{\beta_i\}}, \text{ for } j = \text{L-P, V}; \text{ and } i = 4, 5.$$

Where, j refers to the line and i to the other two situations of risk. β_4 and β_5 were set equal to 1.30 and 2.65 (to achieve, at restriction time, values of survival of 0.57 and 0.63 for the lines V and L-P in $i = 4$ and of 0.2 and 0.25 for the same lines in $i = 5$) mimicking situations of medium and high hazard during the whole life of the

animals (i.e. medium or high hazard associated, for instance, with a farm effect). Afterwards, the mean life restricted to 330 days based on these predicted survival functions was computed, following Klein and Moeschberger (1997).

4.3.3.2.2.- Prolificacy Traits

Three prolificacy traits were studied: Total born (TB), Number born alive (BA) and Number of young alive at weaning (28 days, NW). They were studied using a single-trait repeatability animal model, and the factors included were:

i.- A combination between farm (3 levels: I, II, III) and year-season at the date of the positive pregnancy test (7 levels, changing every two months).

ii.- A combination between genetic type (2 levels: line L-P and line V) and ordinal of parturition (5 levels: 1st, 2nd, 3rd, 4th and $\geq 5^{\text{th}}$).

iii.- Physiological status (2 levels: the doe becomes pregnant at first or she does not).

iv.- A random non additive genetic plus permanent environmental effect, which was assumed to be normally distributed with mean 0 and variance $I\sigma_p^2$.

v.- A random additive genetic effect, which was assumed to be normally distributed with mean 0 and variance $A\sigma_a^2$.

v.- A random residual effect, which was assumed to be normally distributed with mean 0 and variance $I\sigma_e^2$.

The heritabilities and the ratio of the non additive genetic plus permanent environmental variance to the phenotypic variance were those obtained by García and Baselga (2002) for the line V. They were included to take into account the fact that a female could have several records of the trait and that the females were relatives.

In the analysis of these traits all the records from farm I and II were considered but only those from 1st, 2nd and 3rd parturition of farm III. This decision was taken because a very abnormal low prolificacy was observed in the V line at 4th and $\geq 5^{\text{th}}$ parturition that was not observed either in the other farms involved in the experiment, or in females of other lines coetaneous to the V line at the same farm, or in previous generations of the V line.

These traits were analysed by least squares procedures using the program Pest (Groeneveld, 1990).

The F statistic tests reported will be one tailed test, in which the null hypothesis is that the prolificacy between both lines is the same and the alternative is that the V line has a higher prolificacy than the L-P line.

4.3.3.2.3.- Cumulative Production

This trait was defined as the total number of weaned young by a doe during its whole life (from the first positive pregnancy test until its death/culling or censoring time). With this definition, the trait is subjected to right censoring because from live does at the end of the experiment we just know that the real value of this trait is greater than the one recorded until this moment. Thus, we implemented the Bayesian MCMC approach described by Sorensen et al. (1998) to handle normally distributed traits with right censoring.

The model for this analysis included:

- i.- A combination between farm and year-season at first positive pregnancy test, defined every two months (10 levels in total for the combination).
- ii.- Genetic type effect (2 levels: L-P and V lines).
- iii.- An additive genetic effect, which was assumed to be normally distributed with mean 0 and variance $A\sigma_a^2$.
- iv.- A residual effect, which was assumed to be normally distributed with mean 0 and variance $I\sigma_e^2$.

An improper uniform prior was assumed for the covariates effects, a multivariate Normal prior was assumed for the additive genetic effect and they were assumed to be independent. A mass of probability was assigned at points 2401 and 186.8 for the residual and additive genetic variances. These figures were previously estimated with a large data set of the V line described by Sánchez et al. (2005).

The computation of the cumulative production of females in farm III was different to the other farms, due to the problem mentioned above regarding the abnormal prolificacy of the V line at this farm in 4th and $\geq 5^{\text{th}}$ parturitions. Females having more than three parturitions were treated as censored with their cumulative production until this kindling.

After a burn in period of 200,000 rounds and retaining one of each 200 sampled values the convergence of the Markov chain (total length 2,000,000) was assessed by means of the Geweke test for the contrasts of interest.

4.3.3.2.4.- Interval between positive pregnancy tests

We defined this trait as the number of days between two consecutive positive pregnancy tests, and it was considered as a measurement of fertility of the female. This is another example of time until an event trait, where the usage of Survival Analysis techniques is also relevant (Schneider et al., 2005). Moreover, this trait could be subjected to censoring. Thus, if a female dies or is censored (for longevity) after a negative artificial insemination, we just know that the time interval between the last positive pregnancy test and the next one (hypothetical) will be larger than the difference between the death/censoring time and the last positive pregnancy test.

In this case the random censoring assumption does not seem to be fulfilled since when a doe is culled due to infertility it will have a high value for the censoring time which would be associated with a high value for the (unobserved) failure time. However, if we assume an independent censoring schema the likelihood construction follows the same rules as in the case of random censoring (Kalbfleisch and Prentice, 1980). Independent censoring means that an individual can not be censored because of they appear to be at an unusually high or low risk of failure (Kalbfleisch and Prentice, 1980). This is a sign that the failure mechanism acts in the same way over a censored or an uncensored individual. (i.e. A doe culled with the rules given before has the same probability, at censoring time, of having a positive pregnancy test, than another uncensored female (which get a positive pregnancy test at this time) with the same pattern of covariates).

This trait was analysed using a semi-parametric proportional hazard model, similar to the one used in the analysis of the longevity but without frailty term. The estimation procedure was a Bayesian MCMC method. The factors included in the model were:

i.- A time-independent combination between the farm (3 levels: I, II, III) and the year-season at the date of the initiation of the cycle (6 levels, changing every two months).

ii.- A time-independent combination between the cycle order (4 levels: the first from the first positive pregnancy test up to the second, the second from the second up to the third, and so on until the fourth which includes does after their fourth positive pregnancy test) and the genetic type (2 levels: line L-P and line V).

iii.- A time-independent effect of number of young born alive in the parturition that corresponds to the positive pregnancy test which defines the initiation of the interval. This factor was levelled in the same way as for the model to study longevity.

This model did not include either an additive genetic effect or a permanent environmental effect because in the V population, described by Sánchez et al. (2005), the variances associated with these effects were estimated to be very low, thus the effects of these factors can be considered negligible.

An improper uniform prior was assumed for the effect of the covariates.

After a burn in period of 200,000 rounds and retaining one of each 200 sampled values the convergence of the Markov chain (total length 2,000,000) was assessed by means of the Geweke test for the contrasts of interest.

4.3.4.- Results

4.3.4.1.- Functional Longevity

Table 4.3.4.1.1 shows the frequency distribution between censored and uncensored data by genetic type.

Table 4.3.4.1.1.- Count and Proportion of censored/uncensored data by genetic type (L-P and V).

Cell number	L-P	V
Row proportion		
Column proportion		
Censored	179 52.65 67.55	161 47.35 60.98
Uncensored	86 45.50 32.45	103 54.50 39.02

The Chi squared test, correcting for continuity, to check if the distribution of censored and uncensored records under both lines is the same had a value of 2.20, which under the null hypothesis (the same distribution) follows a Chi squared distribution with one degree of freedom. The p-value that corresponds to this statistic was 0.138, so the null hypothesis can not be rejected.

Table 4.3.4.1.2 shows the means, standard deviations, minimums and maximums values for the lives of censored and uncensored records by genetic type.

Table 4.3.4.1.2.- Summary statistics for longevity (in days) of censored/uncensored data by genetic type (L-P and V).

	Mean	L-P	V
S.D. ^a			
Minimum			
Maximum			
Censored	297.9	304.1	
	43.00	44.11	
	204	211	
	393	393	
Uncensored	145.2	146.9	
	84.66	86.27	
	12	7	
	328	330	

^a Standard deviation

A summary on the number and proportion of animals alive at the end of each cycle (defined by the positive pregnancy tests) by genetic type is shown in table 4.3.4.1.3. The period in which more females died or were culled was the interval between the first and the second positive pregnancy test (around 10%), after this period the proportion of dead or culled females decreased.

Table 4.3.4.1.3.- Summary of number and proportion of females alive at each cycle by genetic type (L-P and V) at the end of the experiment.

Cycle order ^a	L-P	V
Initial number (with a first positive pregnancy test)	265 (100%)	264 (100%)
1 st	239 (90.1%)	235 (89.0%)
2 nd	216 (81.5%)	212 (80.3%)
3 rd	198 (74.7%)	191 (72.3%)
4 th	189 (71.3%)	169 (64.1%)
^b 5 th -7 th	179 (67.5%)	161 (60.9%)

^a End of the indicated cycle; ^b Not all the live does have yet had the chance to get a 5th positive pregnancy test

Table 4.3.4.1.4 shows the proportion of females culled or dead out the total of uncensored females within the major culling/death causes. The reproductive disorders have been the main reason of death or culling during the experiment.

Table 4.3.4.1.4.- Proportion of females culled or dead within the major culling/death causes.

Major culling/death causes	Proportion
Reproductive disorders ^a	33.3%
Death ^b	29.2%
Respiratory disorders	11.2%
Digestive disorders	8.5%
Abscess	7.9%
Mastitis	6.3%
Miscellanea	2.6%
Sore hocks	0.5%
Conjunctivitis	0.5%

^a It includes females which died around the parturition date, ^b Females which died in the cage and could not be assigned to a major cause of death.

Table 4.3.4.1.5 shows the marginal posterior statistics for the log-hazard ratios between the two different genetic types within cycle order.

Table 4.3.4.1.5.-Marginal posterior statistics for the log-hazard ratios between genetic types effects (L-P and V) within cycle order (negative values indicate greater risk associated with line V, implying lower longevity).

Cycle order	Contrast	PM ^a	PSD ^b	90HPD ^d	P>0 ^e	PM of RR ^f
1 st	(L-P)-V	-0.18	0.322	-0.716 , 0.347	0.29	0.88
2 nd	(L-P)-V	-0.11	0.342	-0.672 , 0.455	0.37	0.95
3 rd	(L-P)-V	-0.34	0.363	-0.929 , 0.259	0.18	0.76
4 th	(L-P)-V	-1.04	0.439	-1.766 , -0.323	0.007	0.39
≥ 5 th	(L-P)-V	0.26	0.530	-0.565 , 1.191	0.69	1.50

^a Posterior mean, ^b Posterior standard deviation, ^d 90% Highest posterior density region, ^e Probability of a value greater than 0 for the log-hazard ratio, ^f Posterior mean of the relative risk (hazard ratio) between both lines.

The only cycle in which differences for the hazards between lines were clearly different from zero was the fourth, in favour of the L-P line. It was around 2.6 (1.0/0.39) times more likely to cull a V doe or for a V doe to die than a L-P one. In the 5th or greater cycles order difference between the estimates was in favour of the V line but

only 10 L-P and 8 V does died or were culled in these cycles, because many does had not yet reached these cycles.

The posterior mean for the log-hazard ratio between the mean of the L-P and the mean of the V line during the five cycles was -0.28 and its posterior standard deviation 0.208. The 90HPD region for this log-hazard ratio was -0.632 , 0.053, and the probability of having a value greater than 0 was 0.09. The posterior mean of the relative risk (hazard ratio) between both genetic types was 0.77, meaning that to cull a V doe or for a V doe to die was 1.3 times more likely than a L-P one.

The marginal posterior statistics of the restricted mean survival lives for the V and L-P lines, and for the differences between them for the five different situation of risk considered are shown in the table 4.3.4.1.6.

Table 4.3.4.1.6.- Marginal posterior statistics (in days) of the mean survival time restricted to 330 days, by genetic type (L-P and V) and for the contrast between both genetic types, for three levels (low, medium and high) of hazard.

Hazard level	Survival function associated with the hazard level		PM ^a	PSD ^b	90HPD ^c	P>0 ^d
Low	^e $S_{j,1}(t)$	L-P	308.5	14.00	289.4 , 328.0	--
		V	302.0	17.85	276.9 , 326.8	--
		(L-P)-V	6.52	6.573	-2.37 , 16.75	0.92
Medium	^e $S_{j,2}(t)$	L-P	277.3	31.55	230.8 , 323.2	--
		V	258.7	40.66	199.0 , 320.3	--
		(L-P)-V	18.6	18.55	-7.37 , 46.93	0.91
	$S_{j,4}(t) = S_{j,1}(t)^{\exp\{1.30\}}$	L-P	262.5	37.10	206.7 , 318.9	--
		V	246.1	43.00	179.8 , 313.7	--
		(L-P)-V	16.37	14.32	-5.28 , 40.04	0.91
High	^e $S_{j,3}(t)$	L-P	181.3	32.05	128.7 , 234.2	--
		V	150.3	39.28	84.5 , 213.6	--
		(L-P)-V	31.1	31.01	-20.64 , 79.08	0.85
	$S_{j,5}(t) = S_{j,1}(t)^{\exp\{2.65\}}$	L-P	162.3	60.49	62.0 , 258.8	--
		V	140.2	59.26	45.1 , 233.6	--
		(L-P)-V	22.04	19.55	-9.90 , 53.95	0.88

^a Posterior mean, ^b Posterior standard deviation, ^c 90% High posterior density region, ^d Probability of a value greater than 0 for the contrast, ^e Predicted survival functions used to compute life expectancy; the different survival functions were computed assigning the animals combinations of covariates associated with low, medium and high hazard levels.

The marginal posterior distributions were extremely skewed for the low hazard situation, because the real mean survival time restricted to 330 days was close to the upper bound of the parameter space (330 days). In the medium level the skewness was lower and in the high level even lower. The increase in the difference between both lines as we consider situations of higher hazard is due to the fact that the effect of the covariates acts multiplicatively.

4.3.4.2.- Prolificacy traits

Table 4.3.4.2.1 shows the number of records, the crude mean and the crude standard deviation for the three studied prolificacy traits in both lines. This table shows the high prolificacy of both lines.

Table 4.3.4.2.1.- Summary statistics of the prolificacy traits within genetic type (L-P and V).

Trait ^a	L-P			V		
	N ^b	Mean	S.D. ^c	N	Mean	S.D
TB	909	9.6	2.77	890	10.2	2.91
BA	909	8.9	3.24	890	9.4	3.23
NW	843	7.9	2.93	836	8.3	2.80

^a TB = Total born, BA = Born alive, NW = Number of young at weaning, ^b Number of records, ^c Standard deviation.

Table 4.3.4.2.2 shows the contrast between the two different genetic types within parity order for the different prolificacy traits. The V Line showed a better prolificacy performance than L-P line, the differences being significant (p-value < 0.05) between both lines for TB (1st, 2nd, and 3rd parturition), BA (2nd and 3rd parturition) and NW (\geq 5th parturition).

Table 4.3.4.2.2.- Contrast between genetic types(L-P and V) within cycle for prolificacy traits.

Parturition	Contrast	TB ^a			BA ^a			NW ^a		
		Esti. ^b	S.E. ^c	p-value ^d	Esti.	S.E.	p-value	Esti.	S.E.	p-value
1 st	(L-P)-V	-0.55	0.276	0.023	-0.35	0.300	0.119	-0.41	0.264	0.060
2 nd	(L-P)-V	-0.65	0.286	0.011	-0.77	0.312	0.006	-0.44	0.275	0.055
3 rd	(L-P)-V	-0.60	0.297	0.022	-0.72	0.326	0.013	-0.34	0.289	0.118
4 th	(L-P)-V	-0.41	0.421	0.164	-0.14	0.473	0.380	-0.19	0.432	0.330
≥ 5 th	(L-P)-V	-0.20	0.397	0.309	-0.35	0.443	0.209	-0.78	0.467	0.048

^a TB = Total born, BA = Born alive, NW = Number of young at weaning, ^b Estimate, ^c Standard error, ^d p-value in a one tailed test.

Table 4.3.4.2.3 shows the contrast between the two different genetic types for the prolificacy traits along the whole period of the study. These contrasts were computed weighing the differences between both lines within parturition by the proportion of data in each kindling. These contrasts were statistically significant (p-value < 0.05) and in favour of V line for all the traits.

Table 4.3.4.2.3.- Contrast between genetic types (L-P and V) for prolificacy traits.

Trait ^a	Contrast	Estimate	S.E. ^b	p-value ^c
TB	(L-P)-V	-0.53	0.213	0.006
BA	(L-P)-V	-0.53	0.218	0.007
NW	(L-P)-V	-0.41	0.182	0.011

^a TB = Total born, BA = Born alive, NW = Number of young at weaning, ^b Standard error, ^c p-value for a one tailed test.

4.3.4.3.- Cumulative Production

Table 4.3.4.3.1 shows the crude summary statistics by genetic type, for cumulative number of young at weaning for censored and uncensored data.

Table 4.3.4.3.1.- Summary statistics of cumulative production (number of young at weaning) for censored and uncensored data within genetic type (L-P and V).

	L-P		V	
	Censored	Uncensored	Censored	Uncensored
N (%) ^a	185 (69.81%)	80 (30.19%)	166 (62.88%)	98 (37.12%)
Mean	30.0	13.7	32.7	15.4
S.D. ^b	10.31	11.69	10.87	12.48
Min. ^c	4	0	7	0
Max. ^d	65	47	65	50

^a Number of animals and proportion within line, ^b Standard deviation, ^c Minimum, ^d Maximum.

The marginal posterior mean for the contrast between both lines was 5.51 in favour of the L-P line, its posterior standard deviation was 5.979 and the 95HPD region was -6.727 , 16.510. The probability of a value larger than zero was 0.82. Thus, the hypothesis of no differences between both lines cannot be rejected. However, there is some evidence that the L-P line has a larger cumulative number of weaned young than the V line.

4.3.4.4.- Interval between positive pregnancy tests

Table 4.3.4.4.1 shows the summary statistics for censored and uncensored data by genetic type for the interval between two consecutive positive pregnancy tests. During the experimental period the real fertility was 81.2%, which is an acceptable value.

Table 4.3.4.4.1.- Summary statistics of the interval (in days) between positive pregnancy tests for censored and uncensored data within genetic type (L-P and V).

	L-P		V	
	Censored	Uncensored	Censored	Uncensored
N (%) ^a	35 (3.85%)	875 (96.15%)	46 (5.15%)	847 (94.85%)
Mean	89.8	61.9	86.1	61.4
S.D. ^b	27.15	12.72	20.74	12.42
Min. ^c	54	54	57	54
Max. ^d	148	140	141	140

^a Number of animals and proportion within line, ^b Standard deviation, ^c Minimum, ^d Maximum.

The summary of the marginal posterior statistics for the log-hazard ratios between genetic types in every cycle is shown in table 4.3.4.4.2.

Table 4.3.4.4.2.- Marginal posterior statistics for the log-hazard ratios between genetic types (L-P and V) within cycle order (negative values indicate greater risk to get a positive pregnancy test associated with V line, implying higher fertility).

Cycle order	Contrast	PM ^a	PSD ^b	95HPD ^d	P>0 ^e	PM of RR ^f
1 st	(L-P)-V	-0.08	0.095	-0.267 , 0.107	0.19	0.92
2 nd	(L-P)-V	-0.03	0.098	-0.231 , 0.156	0.38	0.97
3 rd	(L-P)-V	-0.04	0.104	-0.238 , 0.170	0.35	0.96
≥ 4 th	(L-P)-V	0.09	0.094	-0.093 , 0.280	0.84	1.10

^a Posterior mean, ^b Posterior standard deviation, ^d 95% High posterior density region, ^e Probability of a value greater than 0 for the log-hazard ratio, ^f Posterior mean of the relative risk (hazard ratio) between both lines.

The posterior mean for the log-hazard ratio between the mean of the L-P line and the mean of the V line during the four cycles was -0.01 and its posterior standard deviation 0.049. The 95HPD region for this log-hazard ratio was -0.106 , 0.084 and the probability of having a value greater than 0 was 0.38. The posterior mean of the relative risk (hazard ratio) between both genetic types was 0.99. All these results indicate the lack of differences in fertility between both lines.

4.3.5.- Discussion and Conclusions

To our knowledge there is only one selection experiment on length of reproductive life in prolific species; it is in mice (Farid et al., 2002). They have selected two lines and have kept a control population. The estimated responses in this experiment after 24 generations were 148 and 105 days (5.3 and 4.4 parturitions) within each line, which supposes between 18% and 22% of the interval between parities per generation. In the French Holstein population the selection of dairy bulls based on their breeding values for functional longevity was implemented in 1997 and recently Ducrocq (2005) has reported a very low but favourable genetic trend which makes sense given the strong selection on milk production and its negative genetic correlation with traits determining functional longevity (fertility, resistance to diseases, etc.).

An adequate method to evaluate the selection procedure used to create the line L-P should have been a comparison with a random sample of does coming from commercial populations in Spain and Portugal. It was not feasible for us to do this and

the comparison has been to the V line, widely spread around the world (Garreau et al., 2004), because its prolificacy and adaptation to different environments. Considering the previous comments, the results have shown that the selection procedure applied here to improve longevity has been successful but not fully expressed through the differences of the restricted mean lives between females of the two lines. However these differences in restricted means could be increased when a higher proportion of data becomes uncensored and unrestricted means can properly be estimated. In the study of the experimental design of our trial we observed that the differences in restricted (to 330 days) lives between two hypothetical lines were only around 40% of the differences between the unrestricted means. The log-hazard ratio between both lines during the experiment was -0.28 and the probability of having a value greater than zero (having higher risk of dying or being culled a L-P female than a V one) was 0.09, thus, given our fixed type I error (0.1), we can reject the null hypothesis of both lines having the same longevity. The difference survival ability is mainly due to differences late in life (fourth cycle), being less important the differences early in life. This difference in survival ability late in life has to be confirmed by keeping this experiment going longer. This result could be expected considering that the selection criterion was a very long life and the genes controlling this trait could be different to those involved in the survival ability early in life. However, some correlated response could be expected since the genetic correlation between early-life (up to second parturition) and later-life (up to third or fourth parturitions) survival ability has been estimated as high (0.8 - 0.9) in other species (dairy cattle: Boettcher et al., 1999; pigs: Tholen et al., 1996). But in our case the selection criterion was very late survival ability, thus it could be expected that the genetic correlation between this trait and early survival ability would be lower.

If we accept that the true difference in mean life restricted to 330 days between the line L-P and the V line is 17 days (corresponding to the medium hazard situation), and that this figure is only a 40 % of the difference for unrestricted lives, the expected value for the difference in unrestricted means would be 42.5 days; this figure is 68.9% of the uncensored interval between parities (61.6 days).

Regarding prolificacy traits the line V showed a better performance, especially for TB and BA, during the first parturitions than the line L-P. However in the 4th and in the 5th parturitions these differences disappeared. Regarding NW significant differences were only detected in favour of the V line in parturitions greater or equal to the fifth (0.7 young). By studying jointly all the parturitions between both genetic types a difference

of 0.53 young in favour of the V line was detected both for TB and BA. The corresponding figure for NW was 0.41 in favour of the V line.

If we admit that line L-P shows better survival ability later in life the results from the functional longevity analysis and those from prolificacy analyses could jointly be explained by the ‘disposable soma’ or the ‘antagonist pleiotropy’ theories of ageing, which postulate an antagonism between prolificacy or reproductive functions (especially at early ages) and longevity (see Kirkwood and Rose (1991) and Kirkwood and Austad (2000) for reviews on these theories explaining the ageing process).

Our results suggest that the L-P line should be tested as a maternal line in the current three ways crossing schema, since no differences have been found in cumulative production of weaned young when this new line has been compared with a highly prolific and very well tested line. Despite the higher prolificacy of the V line, L-P females compensate this disadvantage with a slightly greater longevity and they do not show a lower fertility. It means, that some experiments should be carried out to study, apart from the additive genetic effect of the line L-P for the traits of interest, its combining ability with other lines commonly used in the current three ways crossbreeding schema.

4.4.- References

- Baselga M., 2004. Genetic improvement of meat rabbits, programmes and diffusion. In: Proc. of the 8th World Rabbit Congress. Puebla (Mexico). Vol. 1 pp. 1-13..
- Boettcher P.J., Jairath L.K., Dekkers J.C.M., 1999. Comparison of methods for genetic evaluation of sires for survival of their daughters in the first three lactations. J. Dairy Sci. 82:1034-1044.
- Cifre J., Baselga M., García-Ximénez F., Vicente J.S., 1998a. Performance of a hyperprolific rabbit line I. Litter size traits. J. Anim. Breed. Genet. 115:131-138.
- Cifre J., Baselga M., García-Ximénez F., Vicente J.S., 1998b. Performance of a hyperprolific rabbit line I. Maternal and growth performances. J. Anim. Breed. Genet. 115:139-147.
- Collett D., 2003. Modelling Survival data in medical research, 2nd Edition. Chapman & Hall/CRC Press LLC (USA).
- Damgaard L.H., 2004. Quantitative genetic analysis of survival, linear Gaussian and ordered categorical traits. Ph.D. Thesis. The Royal Veterinary and Agricultural University, Frederiksberg C (Denmark).
- Ducrocq V., 2005. An improved model for the French genetic evaluation of dairy bulls on length of productive life of their daughters. Animal Science 80:249-255.
- Estany, J., Baselga, M., Blasco, A., Camacho, J., 1989. Mixed model methodology for the estimation of genetic response to selection in litter size of rabbits. Livest. Prod. Sci. 21, 67-75.
- Farid A., Crober D.C., Van der Steen H., Patterson D.L., Sabour M.P., 2002. Reproductive performance of mice selected for reproductive longevity. In: Proc. of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier (France), Vol. 30 pp. 681-684.

- García M.L., Baselga M., 2002. Estimation of genetic response to selection in litter size of rabbits using a cryopreserved control population. *Livestock Production Science* 74: 45-53.
- Garreau H., Larzul C., Ducrocq V., 2001. Analyse de la longévité de la souche de lapins INRA 1077. In: Proc. of the 9èmes Journ. Rech. Cunicole, Paris (France). pp. 217 – 220.
- Garreau H., Piles M., Larzul C., Baselga M., de Rochambeau H., 2004. Selection of maternal lines : last results and prospects In: Proc. of the 8th World Rabbit Congress. Puebla (Mexico). Vol. 1 pp. 14-25.
- Groeneveld, E. (1990). Pest user's manual. Institute of Animal Husbandry and Animal Behavior, FAL. Germany.
- Herment A., Runavot J.P., Bidanel J.P., 1994. Une nouvelle évaluation de l'intérêt de la voie hyperprolifique chez le porc. In: Proc. of the 26èmes Journées de la Recherche Porcine en France, 315-320.
- Kirkwood T.B.L., Austad S.N., 2000. Why do we age?. *Nature* 408:233-238.
- Kirkwood T.B.L., Rose M.R., 1991. Evolution of senescence: late survival sacrificed for reproduction. *Phil. Trans. R. Soc. Lond.* 332:15-24.
- Klein J.P., Moeschberger M.L., 1997. Survival analysis techniques for censored and truncated data. Springer-Verlag, New York (USA).
- Noguera J.L., Alfonso L., Babot D., Pérez-Enciso M., Estany J., 1997. Resultados de un experimento de selección del tamaño de camada mediante un esquema hiperprolífico en porcino. VII Jornadas sobre Producción Animal, ITEA Vol. Extra No. 18(1):391-393.
- Ramón J., Rafel O., 2002. 1991-2000. Diez años de gestión global en España. In: Proc. of Expoaviga 2002, X jornada cunícola, Barcelona (Spain), pp.113–117.

Richard M., David P.J., 1985. Effectiveness of genetic selection for prolificacy in pigs. J. Reprod. Fert. Suppl. 33:127-138.

Sánchez J.P., Korsgaard I.R., Damgaard L.H., Baselga M., 2005. Analysis of rabbit does longevity using a semiparametric log-Normal animal frailty model with time-dependent covariates. Submitted to Genet. Sel. Evol.

Schneider M.D., 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.

Sorensen D., Vernersen A.H., 1991. Large scale selection for number of born piglets using an animal model, In: Proc. 42nd Annual meeting of the EAAP, Commission on Animal Genetics, Session 3. Berlin.

Sorensen D.A., Gianola D., Korsgaard I.R., 1998. Bayesian mixed-effects models analysis of a censored normal distribution with animal breeding applications. Acta Agric. Scand., Sect. A, Animal Sci. 48:222-229.

Tarres J., Piedrafita J., Ducrocq V., 2005. Validation of an approximate approach to compute genetic correlations between longevity and linear traits. Genet. Sel. Evol. (*in press*).

Tholen E., Bunter K.L., Hermes S., Graser H-U., 1996. The genetic foundation of fitness and reproduction traits in Australian pig populations 2. Relationships between weaning to conception interval, farrowing interval, stayability, and other common reproduction and production traits. Aust. J. Agric. Res. 47:1275-1290.

Tudela F., Hurtaud J., Garreau H., de Rochambeau H., 2003. Comparaison des performances zootechniques de femelles parentales issues d'une souche témoin et d'une souche sélectionnée pour la productivité numérique. In: Proc. of the 10^{èmes} Journ. Rech. Cunicole, Paris (France) pp. 53 – 56.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

CHAPTER 5: DISCUSSION AND CONCLUSIONS

In the definition of the model to genetically study longevity in a population, using proportional hazard models, the first step is to decide if a parametric model is suitable for describing the ageing process, or a semiparametric model is preferable. Most of the time this choice is based on computational aspects, because the semiparametric models have such high computational demands, which makes it unfeasible to analyse huge data sets. In rabbit populations we are not limited due to the size of the data sets, thus, slight violations of the requirements of a particular parametric model could be enough to choose a semiparametric model. In this thesis we initially defined the length of productive life of the females as the number of days since the date of the first presentation to the buck to the death/culling or censoring date. For this definition we study the adequacy of the Weibull model to fit the data throughout a graphical test of the Kaplan-Meier estimator. We observed that at around 31 days (first parturition) the fitting failed. A second definition of the trait also has been used: days since the first positive pregnancy test. In this case we used the same graphical test, but based on the estimated baseline hazard function, to decide if a Weibull model properly fits the baseline hazard function. In this second case the result from this graphical test was even worse than for the previous definition. Thus we decided to use semiparametric models. Previous studies in rabbits (Garreau et al., 2001) used the grouped data model (Prentice and Gloeckler, 1978) to fit longevity in rabbit females, measured as the number of inseminations that a doe has during its whole life. In pigs, Yazdi et al. (2000b) also studied longevity using survival analysis techniques. They used the same graphical test as us in the first case to study the suitability of the Weibull model. They observed a slight misfit after the first weaning, where there was an intensive culling. But, in spite of this misfit they preferred to use it.

In our case, two more definitions of the trait could have been useful to allow for the usage of the Weibull model, reducing the computation requirements. The first one, days from first parity until death or culling, however with this definition a lot of information would be lost since a high proportion of females die between the first presentation to the buck or the first positive pregnancy test and the first parity. The second one, using a similar procedure to the one implemented in dairy cattle by Roxstöm et al. (2003), days from the first positive pregnancy test until death or culling

but defined in a cycle basis, and allowing for different baseline hazard functions in each cycle.

With the first definition of the trait all the main effects proposed to explain longevity (year-season, number born alive every parturition, reproductive rhythm, ordinal of cycle and age of the female at the first presentation to the buck) showed a significant effect. Moreover the interactions between number born alive and reproductive rhythm and between ordinal of cycle and reproductive rhythm also showed a significant effect. In general, those levels of systematic effects which are, in some sense indicators of sick does (i.e. very low number of born alive) were the ones with higher associated risk. This result had also been previously reported by Garreau et al. (2001). Regarding the reproductive rhythm, it was observed that does delayed at first parturition had a lower survival probability than non delayed females during their whole lives. A doe delayed at second parturition showed a lower risk of dying/being culled than a non delayed female, when the comparison is done at time of the parturition of the latter. This is explained by the high risk associated with the parturition, especially at these ages, when the does are still growing. If the comparison is done at the parturition date of the former the differences disappear. Delays at late ages had a negative but negligible effect. With the first definition of the trait, the age of the does at date of the first presentation to the buck also showed a significant effect - a doe was more likely to die or be culled as the start of its productive life was earlier. When the longevity was defined from the first positive pregnancy test, very similar conclusions regarding the systematic effects were found. But in this case the age of the female at the beginning of its productive life did not show a significant effect. When the second definition of the trait was used, the reproductive rhythm of the female was defined in a more precise way. It was observed that empty does were those subjected to a higher risk, followed by pregnant, lactating and finally those does in which there was overlapping between two consecutive cycles. Again, it can be concluded that those situations relating to sick does (i.e. very low prolificacy or infertility) had high associated risk.

The estimated heritability on the log-g(t) scale with the first definition of the trait was 0.053, but it was 0.095 with the second definition. This improvement could be due to a better fitting of the physiological status of the female during its whole life, and to the year-season effect which was defined every three months instead of every six months and because an animal model was used. The second estimated value for the heritability on log-g(t), which is the most reliable, is larger than the most reliable value

reported by Garreau et al. (2001), but is in the range of previous estimates for this parameter in other species obtained using survival analysis techniques (dairy cattle: Chirinos, 2003 (Revision there in); pigs: Yazdi et al., 2000a,b). In beef cattle the previous estimated values are lower than our estimate (Diaz et al., 2002; Hyde et al., 2000).

With the second definition of the trait the model also includes a residual term in log-frailty. It seems desirable to include this effect in order to account for further individual heterogeneity, other than the additive genetic. However, when the estimated heritability was compared with that from a model without this residual term a nearly identical figure was obtained. This could be consequence of an increase in the scale of the model with residual in log-frailty but more research is needed in order to clarify this possible scale effect. However, Damgaard et al. (2004) showed that when this individual heterogeneity is ignored some bias could exist related to the censoring level.

Regarding the estimation of correlations between longevity (defined using the second approach) and prolificacy traits, neither the number of young alive at parturition nor the number of young alive at weaning showed genetic correlations statistically different from zero with functional longevity. This result is in agreement with a previous study (Theilgaard et al. 2005) in which no correlated response on longevity was found after selection for prolificacy. Between the number of young alive at weaning and functional longevity a not significantly different from zero environmental correlation was found. Between the number of young born alive at parturition and functional longevity a very low favourable (-0.11 ± 0.016) and statistically different from zero environmental correlation was found, indicating that some environmental factors act in the same way over functional longevity and over the number of young alive at parturition. No previous estimations of the correlations between functional longevity and prolificacy traits have been found in rabbit literature. In other species the statistical techniques employed to infer these correlations did not allow for survival analysis methods to study longevity. The reported values in pigs indicate a medium-low genetic correlation between prolificacy and longevity, being in most cases favourable (Tholen et al., 1996; Serenius and Stalder, 2004). Only Tholen et al. (1996) found in one of the studied populations a negative genetic correlation between the number of piglets born alive and several measures of stayability.

With respect to the evaluation of the L-P line the results are within the expected. It seems that the L-P line shows better survival ability than the V line, especially later in

life (fourth cycle). Moreover if both lines are compared during the whole studied period a log-hazard of -0.28 was detected, which for our type I error (0.1) can be considered as significant. If this result is expressed in time units using the differences in restricted (to 330 days) expectancy of life between lines, the figures obtained depend on the level of hazard associated with other factors in the model. However, for a medium hazard level the difference in mean restricted life is around 17 days. This difference would correspond to 42 days in unrestricted life expectancy. This value is 68.9% of the kindling interval in our experiment. The only selection experiment on reproductive longevity in prolific species has been carried out in mice (Farid et al., 2002). In this experiment the estimated responses by generation ranges between 18 and 22% of the kindling interval (depending on the studied line), 4.4 and 6.2 days per generation in absolute value. Regarding prolificacy traits line V showed a better performance than the L-P line, especially early in life, the mean differences during the whole experimental period were 0.53 young for total born and born alive and 0.41 for number at weaning. The results from longevity and prolificacy analyses could jointly be explained within the frameworks of the 'disposable soma' or the 'antagonist pleiotropy' theories of ageing (Kirkwood and Austad, 2000). Based on the performances on the L-P line, no differences have been found in cumulative production of weaned young when this new line has been compared with a highly prolific and very well tested line, this new line could be a candidate to be included in the current three ways crossing schema of rabbit production. However, before taking this decision some experiments should be carried out to study, apart from the additive genetic effect of the line L-P for the traits of interest, its combining ability with the other lines commonly used in the current three ways crossbreeding schema.

5.1.- References

- Chirinos Z.R., 2003. Desarrollo de un sistema de valoración genética para la longevidad en ganado frisón español. Tesis doctoral. Universidad Politécnica de Madrid (Spain).
- Damgaard L.H., 2004. Quantitative genetic analysis of survival, linear Gaussian and ordered categorical traits. Ph.D. Thesis. The Royal Veterinary and Agricultural University, Frederiksberg C (Denmark).

- Diaz C., Chirinos Z., Moreno A., Carabaño M.J., 2002. Preliminary analysis of functional longevity in the Avileña Negra Ibérica beef cattle breed. In: Proc. of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier (France), Vol. 29 pp. 697-700.
- Essl A., 1998. Longevity in dairy cattle breeding: a review. *Livestock Production Science* 57:79-89.
- Farid A., Crober D.C., Van der Steen H., Patterson D.L., Sabour M.P., 2002. Reproductive performance of mice selected for reproductive longevity. In: Proc. of the 7th World Congress on Genetics Applied to Livestock Production, Montpellier (France), Vol. 30 pp. 681-684.
- Garreau H., Larzul C., Ducrocq V., 2001 Analyse de la longévité de la souche de lapins INRA 1077. In: Proc. of the 9^{èmes} Journ. Rech. Cunicole, Paris (France). pp. 217 – 220.
- Hyde L.R., 2000. An application of survival analysis to the genetic evaluation of reproductive life of beef females. Ph.D. Thesis. Colorado State University, Fort Collins, Colorado (USA).
- Kirkwood T.B.L., Austad S.N., 2000. Why do we age?. *Nature* 408:233-238.
- Prentice R., Gloeckler L., 1978. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 34:57-67.
- Roxstöm A., Ducrocq V., Strandberg E., 2003. Survival analysis of longevity in dairy cattle on a lactation basis. *Genet. Sel. Evol.* 35:305-318.
- Serenius T., Stalder K.J., 2004. Genetics of length of productive life and lifetime prolificacy in the Finnish Landrace and Large White pig populations. *J. Anim. Sci.* 82:3111-3117.

Theilgaard P., Sánchez J.P., Pascual J.J., Friggens N.C., Baselga M., 2005. Effect of selection for prolificacy, and body fatness on survival of rabbit does assessed using a cryopreserved control population. Submitted to *Livestock Production Science*.

Tholen E., Bunter K.L., Hermesch S., Graser H-U., 1996. The genetic foundation of fitness and reproduction traits in Australian pig populations 2. Relationships between weaning to conception interval, farrowing interval, stayability, and other common reproduction and production traits. *Aust. J. Agric. Res.* 47:1275-1290.

Yazdi M. H., Lundeheim, N., Rydhmer L., Ringmar-Cederberg E., Johansson K., 2000a. Survival of Swedish Landrace and Yorkshire sows in relation to osteochondrosis: a genetic study. *Animal Science* 71:1-9.

Yazdi M. H., Rydhmer L., Ringmar-Cederberg E., Lundeheim, N., Johansson K., 2000b. Genetic study of longevity in Swedish Landrace sows. *Livestock Production Science* 63, 255 – 264.