

ANÁLISIS DE SUPERVIVENCIA EN LA EVALUACIÓN GENÉTICA DE *VIDA PRODUCTIVA* EN RODEOS LECHEROS: UNA INTRODUCCIÓN

*Survival analysis for the genetic evaluation of productive life of dairy
herds: a primer*

**SCHNEIDER, M. DEL P.¹, CANTET, R. J. C.^{2,3}, SANTOS CRISTAL,
M.G.²**

1. Cátedra de Producción Animal II. Facultad de Ciencias Agrarias, Universidad Nacional del Litoral. P. Kreder
2805, 3080 Esperanza. Argentina.
2. Departamento de Producción Animal. Facultad de Agronomía, Universidad de Buenos Aires, Argentina
3. Consejo Nacional de Investigaciones Científicas y Técnicas. Argentina.

RESUMEN

Las evaluaciones genéticas para el carácter *vida productiva* requieren emplear métodos y modelos estadísticos apropiados que contemplen que: a) cuando se realiza la evaluación, algunos animales aún están produciendo; en consecuencia, su *vida productiva* es desconocida, y b) los efectos que afectan la *vida productiva* no actúan de modo lineal, pudiendo variar en el tiempo. El análisis de supervivencia es una herramienta estadística apropiada para analizar *vida productiva*. Se basa en modelar una *función de riesgo* $\lambda(t)$, la cual describe la probabilidad de eliminación de

un animal en el momento t , condicional a que el animal sobrevive hasta dicho tiempo. La *función de riesgo* para un animal es igual al producto de una *función de riesgo base* ($\lambda_0(t)$), que describe el proceso de envejecimiento, y un vector que contiene los efectos fijos y aleatorios que afectan el proceso de descarte, los cuales pueden depender del tiempo. La *función de riesgo* obtenida a partir de la distribución Weibull permite un análisis estadísticamente informativo para *vida productiva* con modelos complejos y bases de datos de gran tamaño. En la actualidad, países como Canadá, Francia, Alemania e Italia realizan evaluaciones genéticas de *vida productiva* empleando modelos de padre o de padre y abuelo materno, asumiendo una distribución de riesgo Weibull y empleando la metodología del análisis de supervivencia. En esta presentación se realiza una descripción introductoria al empleo de dicha metodología estadística, en la predicción del valor de cría de toros lecheros para el carácter *vida productiva*.

PALABRAS CLAVE: *análisis de supervivencia, vida productiva, toros lecheros.*

SUMMARY

Genetic evaluation for longevity traits requires proper methods and statistical models as: a) some animals are still in production when genetic evaluations are produced, hence the length of productive life is unknown, and b) the effects that influence the length of productive life do not act linearly, and they even may change with time. Survival Analysis is an appropriate statistical tool to analyze longevity. The models used are based on a *hazard function* $\lambda(t)$ which describes the risk of being culled at time t , given that the animal had survived to time t . The *hazard function* for an animal is defined as the product of the *baseline hazard function* ($\lambda_0(t)$), which describes the aging process, and a vector of variables that influence the culling process which, in turn, can be time dependent. The *hazard function* assuming a Weibull distribution provides an

informative statistical analysis of *productive life* for complex models and for large data sets. At present time, countries such as Canada, France, Germany and Italy produce genetic evaluations for *productive life* using sire or sire-maternal grand sire models, assuming a Weibull density by means of survival analysis methodology. This presentation introduces the use of survival analysis methodology for the genetic evaluation of dairy bulls, based on the *productive life* of their daughters.

Keywords: survival anaalysis, productive life, dairy bulls.

INTRODUCCIÓN

En la producción animal, especialmente en la producción lechera, la vida productiva es una característica importante que afecta la rentabilidad de la empresa agropecuaria. El incremento de la vida productiva de una vaca lechera tiene como consecuencias: a) una reducción en la cantidad de vaquillonas de reposición que se deben criar; b) un incremento en la producción media del rodeo, a través de un aumento en la proporción de vacas adultas de mayor nivel de producción; c) una mayor disponibilidad de recursos que se pueden utilizar para aumentar el tamaño del rodeo en ordeño y d) un aumento en la intensidad de selección para las características de producción.

Kachman (1998) reseñó los problemas que existen con los datos al realizar una evaluación genética para vida productiva (medida en días, meses o años). Entre ellos se puede citar: 1) definir el momento inicial y final para el intervalo de medición; 2) definir cómo considerar datos provenientes de vacas que dejan el rodeo por factores no relacionados con la producción (por ejemplo, ventas); 3) definir cómo tratar los registros de aquellas vacas que aún se encuentran productivas al momento de la evaluación (datos censurados); 4) considerar que la distribución de vida productiva es *asimétrica* (mayor frecuencia de vacas con menor vida

productiva) . Un enfoque pertinente para analizar este carácter es el análisis de supervivencia (*survival analysis*), metodología estadística originalmente desarrollada para tratar investigaciones en medicina e ingeniería. En la actualidad, se lo emplea para seleccionar animales con mayor mérito genético por vida productiva, en la evaluación genética de bovinos de leche (Ducrocq y Solkner, 1998b; Harbers, 1999), aves (Ducrocq, Besbes y Protais, 1999), porcinos (Ringmar-Cederberg, Johansson, Lundeheim y Rydhmers, 1997) y caballos de salto (Ricard y Fournet-Hanocq, 1997). El objetivo de esta presentación es describir de modo introductorio, la metodología del análisis de supervivencia en las evaluaciones genéticas de bovinos de leche para *vida productiva*.

2. VIDA PRODUCTIVA Y ANÁLISIS DE SUPERVIVENCIA

Ducrocq (1987) definió como *vida productiva verdadera*, la aptitud de una vaca para permanecer en producción dentro de un rodeo, cuando no se toman en cuenta las razones por las cuales pueda ser eliminada. Mide además, la percepción que tiene el tamero sobre el valor de la vaca (Ducrocq, Quaas, Pollak y Casella, 1988a). Ducrocq y otros (1988a) agregaron que es importante distinguir entre el refugio *involuntario* (ejemplo, venta de una vaca que no se preñó pero rentable), del refugio *voluntario* (ejemplo, venta de una vaca en buen estado de salud pero no rentable). Por lo tanto, Ducrocq, Quaas, Pollak y Casella (1988b) definieron *vida productiva funcional* a la capacidad del animal para retrasar el refugio involuntario, es decir, para permanecer en el rodeo sin enfermarse ni tener problemas reproductivos. Esta variable es ajustada por el nivel de producción de leche (Boettcher, Jairath y Dekker, 1999). Desde el punto de vista del mejoramiento genético animal, es de particular interés la vida productiva funcional porque tiene en cuenta el nivel de producción, principal causa de eliminación de vacas en los tambos, de forma tal que se puedan detectar diferencias genéticas entre animales para las

distintas componentes de la producción, como por ejemplo, salud, fertilidad, conformación, vejez, etc. En los últimos años, mejorar la vida productiva funcional en los rodeos lecheros cobró importancia, dado el incremento en los costos de reposición de vientres y la intensa selección por producción.

La vida productiva es un carácter difícil de medir debido a la presencia de datos censurados en aquellas hembras que aún siguen produciendo al momento del análisis, con lo cual sólo se puede estimar el límite inferior de su vida productiva. Se trató de resolver este problema utilizando indicadores sencillos de vida productiva, como por ejemplo la *permanencia productiva (stayability)*, índice que establece si una vaca todavía está viva o muerta (carácter discreto binario) en un momento dado, cuya definición es variable: Everett y otros. (1976) consideraron los 54 meses; Schaeffer y Burnside (1974), el número de lactaciones, y VanRaden y Klaaskate (1993) tomaron una proyección del número de lactancias sólo para las vacas que continúan en producción. El método consiste en una serie de ecuaciones de predicción de la vida productiva en base a la edad de la vaca al momento de la evaluación, su edad a la primera parición, la duración de la lactación y si está o no lactando. Al emplear las definiciones de *vida productiva funcional* sugeridas por Everett, Keown y Clapp. (1976) y Schaeffer y Burnside (1974), se produce una pérdida de información. La limitación del método de VanRaden y Klaaskate (1993) se debe a que el reducido número de registros disponibles impide predecir la vida productiva de una vaca con una precisión aceptable (Ducrocq, 1997).

La predicción de valores de cría empleando métodos lineales tradicionales no es apropiada, debido a que el carácter *permanencia productiva* posee una distribución binomial. En consecuencia, se emplea el análisis estadístico de supervivencia que puede combinar la información de *observaciones no censuradas* (vacas que finalizaron su vida productiva) y *censuradas* (vacas todavía en producción al momento de la evaluación genética). Además, el

análisis de supervivencia permite considerar covariables que son función del tiempo y afectan la vida productiva: nivel de producción de leche y/o sólidos, cambios anuales en el tamaño del rodeo, estado sanitario y manejo. El análisis de supervivencia en la evaluación genética animal emplea grandes bases de datos y modelos de análisis complejos (Smith y Quaas, 1984; Ducrocq, 1987; Ducrocq, 1994; Dürr 1999; Vukasinovic, Moll y Künzi, 1997; Schneider, 1998; Vollema y Groen, 1998; Jong, Vollema, Van der Beek y Harbers, 1999; Pasman y Reinhard, 1999; Schneider y Miglior, 1999). En una de las primeras investigaciones realizadas empleando esta metodología, Famula (1981) asumió que la supervivencia de vacas lecheras era una variable aleatoria exponencial afectada por ciertas covariables y discutió las ventajas de utilizar observaciones censuradas. Smith y Quaas (1984) emplearon el análisis de supervivencia para predecir los valores de cría de toros lecheros con mediciones de la vida productiva de sus hijas. Ducrocq (1994) y Ducrocq y Sölkner (1998a) desarrollaron un programa de cálculo para ajustar modelos mediante el análisis de supervivencia, que puede usarse en aplicaciones de gran escala. En la actualidad, países como Francia (Ducrocq, 1999) Austria (Ducrocq y Sölkner, 1998b), Alemania (Pasma y Reinhardt, 1999) y Holanda (Harbers, 1999), utilizan el análisis de supervivencia en las evaluaciones genéticas para predecir el valor de cría por vida productiva de toros Holstein y de otras razas de menor difusión. Otros países europeos (Harbers, 1999; Pasman y Reinhardt, 1999; Strandberg y Roth, 1999; Schneider y Miglior, 1999) y Canadá (Dürr, 1999; Schneider, 1998) probaron esta metodología para incorporarla a sus evaluaciones genéticas rutinarias.

3. METODOLOGÍA DEL ANÁLISIS DE SUPERVIVENCIA

El análisis de supervivencia permite estudiar la ocurrencia y el momento de ciertos eventos o sucesos (Allison, 1995), empleando modelos causales o predictivos donde la

ocurrencia de un evento es función de distintas variables. Ejemplos en las ciencias sociales y en las naturales corresponden a la aparición de una enfermedad, la ocurrencia de fallas en maquinarias, colapsos de los mercados financieros, el incremento en el número de arrestos o divorcios, vida productiva de animales, etc. La ventaja del análisis de supervivencia consiste en ajustar más eficientemente que los métodos estadísticos tradicionales, caracteres censurados y efectos fijos que son función del tiempo.

3.1. Tiempo de supervivencia (*survival time*)

El *tiempo de supervivencia (survival time)* es el que transcurre hasta la aparición de un cierto evento (Allison, 1995), el cual se denomina *falla (failure)*, porque se refiere a experiencias negativas tales como el desarrollo de una enfermedad, la eliminación de una vaca en un rodeo, la muerte (Kleinbaum, 1996; página 4). No obstante el *tiempo de supervivencia* puede corresponder, por ejemplo, al tiempo necesario para volver al trabajo luego de una operación, con lo cual el evento o *falla* tiene una connotación positiva. Entonces, el *tiempo de supervivencia* es una variable aleatoria positiva que mide la distancia o el largo de un intervalo entre un punto de origen (cuando comienza la recolección de datos) y el evento de interés o *falla*. De este modo, el tiempo de supervivencia no sólo se refiere a un número de días (o meses, o años), sino que también puede expresar el número de kilómetros recorridos por un auto antes del reemplazo de las cubiertas, o el número de rotaciones de un motor antes de su rotura, o la cantidad de kilos de leche producidos por una vaca durante sus años en el rodeo, etc. El carácter *vida productiva*, en el contexto de la evaluación genética animal, se define como el número de días que transcurren entre el primer parto y el descarte del animal (Ducrocq y Solkner, 1988b).

3.2. Censurado

Uno de los aspectos distintivos del análisis de supervivencia dentro de las ciencias biológicas, es que ciertos individuos estudiados no experimentan el evento de interés o *falla*, al terminar el período de análisis (Kleinbaum, 1996; páginas 5 a 7). Por ejemplo, cuando algunas vacas todavía están en producción y no han sido descartadas, siendo su *tiempo de supervivencia* desconocido. Estas observaciones se llaman *censuradas* (*censored records*) y su ocurrencia se denomina *censurado* (*censoring*). Existen dos tipos de censurado: *por la derecha* (*right censoring*) y *por la izquierda* (*left censoring*). El censurado por la izquierda ocurre cuando el evento no se puede registrar porque ocurrió previo al período de observación. En general no suele ser de interés en las aplicaciones al mejoramiento genético, porque aquellos animales que mueren antes del comienzo del período de observación no son incluidos en el análisis. El censurado por la derecha ocurre cuando el evento no es observado debido a que la recolección de datos fue discontinuada. En el ganado lechero interesa el censurado por la derecha, que se produce cuando: 1) una vaca se encuentra en producción al final del período de análisis; o 2) el tambo abandona el control lechero oficial, o 3) se vende una vaca que luego no puede ser ubicada en la base de datos (Ducrocq, 1987).

3.3. Funciones estocásticas del tiempo de supervivencia

El tiempo (t_i) durante el cual se produce la *falla* (descarte) del animal i se considera un proceso aleatorio, función de efectos que se combinan en un factor de riesgo: $n_i = x_i \beta + z_i \mu$, donde β son los efectos fijos de sexo, tambo, etc, y μ es el valor de cría o mérito genético del animal. Los elementos x_i y z_i son las filas de las matrices de incidencia de los efectos fijos y los valores de cría, respectivamente. La *función de supervivencia* $S(t)$ es la probabilidad de que un animal siga vivo en el momento t . Alternativamente, $S(t)$ puede considerarse como la proporción

de animales que aún permanecen vivos en el momento t . La función de distribución acumulada $F(t)$ hasta el tiempo t , mide la probabilidad de que un individuo *falle* (sea descartado) antes del tiempo t . Las funciones $S(t)$ y $F(t)$ se relacionan del siguiente modo:

$$S(t) = \Pr(T \geq t) = 1 - F(t) \quad 0 < t < \infty$$

En la expresión anterior, T es la variable aleatoria y t es un valor fijo en el espacio paramétrico de la variable T . Finalmente, la *función de riesgo* ($\lambda(t)$, *hazard function*) es la *probabilidad condicional* de que un individuo muera en el lapso $[t, t + \Delta t]$, dado que sobrevivió hasta el tiempo t :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob} [t \leq T < t + \Delta t \mid T \geq t]}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{\partial \log S(t)}{\partial t}$$

La *función de riesgo* $\lambda(t)$ mide la posibilidad de descartar una vaca por unidad de tiempo, dado que la hembra se mantuvo en producción hasta el tiempo t . Nótese que, en contraste con la *función de supervivencia* $S(t)$ que enfatiza el hecho de no descartar la vaca (que no *falle*), la *función de riesgo* $\lambda(t)$ focaliza la ocurrencia del descarte (*falla*). Es decir que, cuanto mayor es el valor que toma $S(t)$ menor es $\lambda(t)$, y viceversa. Como muestra la expresión superior, la *función de riesgo* $\lambda(t)$ no es una probabilidad, sino un cociente entre dos cantidades. El numerador es la probabilidad condicional de que el descarte de la vaca ocurra en el intervalo de tiempo entre t y $t + \Delta t$, dado que el tiempo de supervivencia T es mayor o igual a t . El denominador, Δt , indica un pequeño intervalo de tiempo. La escala para este cociente va de 0 a ∞ y depende de la unidad de medida del tiempo (días, semanas, meses, años). Un valor de la *función de riesgo* $\lambda(t)$, a un

tiempo t determinado, se define como la *tasa de riesgo (hazard rate)*, para el evento o *falla* descarte de la vaca.

3.4. Distribuciones de supervivencia

La incertidumbre asociada con la ocurrencia de *fallas* puede ser modelada empleando distribuciones estadísticas tales como la Exponencial, la Weibull, la Lognormal, la Gamma, la Log-logística y la Gompertz. La densidad Weibull es la más empleada para modelar la distribución de la vida productiva de vacas lecheras: es una generalización de la exponencial, de la cual difiere porque el riesgo (*hazard rate*) no toma un valor constante, siendo en consecuencia más flexible para el análisis. La función de densidad de la Weibull es igual a:

$$f(t) = \lambda \rho t^{(\rho-1)} \exp(-\lambda t^\rho),$$

El parámetro ρ determina la forma de la densidad, mientras que λ se asocia con la escala. El riesgo aumenta cuando $\rho > 1$, disminuye cuando $\rho < 1$ y es constante cuando $\rho = 1$. En este último caso la distribución es igual a una exponencial con parámetro λ . El asumir una tasa de riesgo constante implica, por ejemplo, que la probabilidad de que un animal sobreviva cinco años más es igual al momento del nacimiento, o los 5 años de edad, o a los 10 (Kachman, 1998), suposición que es muy poco realista en vacas lecheras.

3.5. Modelo

El análisis de supervivencia se realiza mediante el siguiente modelo para la *función de riesgo* $\lambda(t;\mathbf{x})$:

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

El primer término ($\lambda_0(t)$) es común para todos los animales y se denomina *función de riesgo base*

(*baseline hazard function*), mientras que el segundo ($\exp\{\mathbf{x}'\boldsymbol{\beta}\}$) considera en el vector de parámetros $\boldsymbol{\beta}$, los efectos fijos del proceso de descarte, tales como tambo o estado de la lactancia. La función de riesgo base tiene una forma paramétrica, asumiendo una distribución exponencial, o una Weibull, o puede tener una forma semiparamétrica más general (modelo de Cox, 1972).

3.6. Estimación

La estimación de los efectos de los efectos fijos en la curva de supervivencia se suele realizar mediante métodos basados en la función de verosimilitud, diferenciándose según los modelos sean paramétricos o semiparamétricos. En el primer caso, la contribución a la verosimilitud de una observación no censurada es la *función de densidad del tiempo de falla*, mientras que para una observación censurada es la *función de supervivencia al momento del censurado* (Ducrocq, 1987). La función de verosimilitud completa se obtiene multiplicando las contribuciones de todos los datos. La estimación de los parámetros se realiza maximizando el logaritmo de la verosimilitud resultante (Ducrocq, 1997). Los parámetros de la *función de riesgo base* y de $\boldsymbol{\beta}$ son estimados conjuntamente.

Para modelos semiparamétricos, Cox (1972) sugirió estimar $\boldsymbol{\beta}$ mediante la maximización de una verosimilitud donde $\lambda_0(t)$ se halla marginalizada, denominada *verosimilitud parcial* (*partial likelihood*). De este modo, dicha función depende sólo de $\boldsymbol{\beta}$. Una vez obtenidas las estimaciones de estos parámetros, se estima la *función de riesgo base* empleando una aproximación similar a la utilizada para calcular los estimadores Kaplan-Meier (Ducrocq, 1997).

3.7. Efectos dependientes del tiempo

Efectos tales como el sexo, la raza, el rodeo y la edad al primer parto, son independientes del tiempo. Sin embargo, existen variables que se encuentran afectadas por el tiempo, por ejemplo los años, las estaciones, los cambios en el tamaño de los rodeos, el número de lactaciones, u otras que afectan el momento de eliminación de la vaca (*failure time*). Los modelos de *riesgo proporcional* incluyen situaciones en las cuales efectos en \mathbf{x} varían con el tiempo: $\mathbf{x}(t)$. En tal caso, el modelo debe modificarse del siguiente modo:

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

En este modelo el tiempo es discretizado en intervalos dentro de los cuales se asume un *riesgo proporcional*, que cambia de un intervalo a otro. Los efectos fijos que dependen del tiempo pueden tener varios niveles, que se asocian con cada observación de vida productiva. Por ejemplo, la vida productiva de cada vaca se relaciona con los distintos niveles del efecto año de lactación que le correspondan.

3.8. Modelos frágiles

En las evaluaciones genéticas animales, los modelos de *riesgo proporcional* incluyen efectos aleatorios, como en los clásicos modelos lineales mixtos, cuando los caracteres tienen distribución normal. En el análisis de supervivencia los modelos mixtos son conocidos como modelos *frágiles* (*frailty models*). En dichos modelos, se incluye la variable aleatoria no observable \mathbf{v} , para un individuo, o \mathbf{v}_q para un grupo de individuos (por ejemplo, todas las hijas de un toro), afectando multiplicativamente al riesgo. Estas variables describen las características genéticas compartidas no observables que actúan sobre el riesgo de cada hija (Ducrocq, 1997). Así, para el individuo i podemos definir:

$$\mathbf{W}'_i = [\mathbf{x}'_i : \mathbf{z}'_i] \quad \boldsymbol{\theta}' = [\boldsymbol{\beta}' : \mathbf{u}']$$

siendo \mathbf{x}'_i y \mathbf{z}'_i filas de las matrices de incidencia que relacionan la observación de i , con los vectores de parámetros $\boldsymbol{\beta}$ y de variables aleatorias \mathbf{u} . Se suele asumir que todos los efectos son tiempo-dependientes, con una única *función de riesgo base* y un solo efecto aleatorio (por ejemplo, la mitad del valor de cría del padre). Al combinarse estos efectos en el modelo Weibull, la función de riesgo $\lambda(t)$ para el animal i es igual a:

$$\lambda(t | \boldsymbol{\theta}, \rho) = \rho t^{(\rho-1)} \exp\{\mathbf{W}'_i \boldsymbol{\theta}\} = \rho t^{(\rho-1)} \exp\{\rho \log \lambda + \mathbf{W}'_i \boldsymbol{\theta}\}$$

En el caso de que las observaciones pertenezcan a la hija i de un toro q , con un *tiempo para que ocurra la falla (failure time) T_i* , tenemos el siguiente modelo:

$$\lambda(t | \boldsymbol{\theta}, \rho) = \rho t^{(\rho-1)} \nu_q \exp\{\mathbf{x}'_i \boldsymbol{\beta}\} \quad \text{para } t \leq T_i$$

donde $\nu_q = e^{s_q}$ es el *modelo frágil* y s_q es la mitad del valor de cría del toro q .

Debido a su simplicidad, la distribución gamma es empleada para modelar la variabilidad del término ν (Ducrocq y otros, 1988b), aunque no posee las propiedades convenientes de la distribución normal para describir un modelo poligénico infinitesimal (Ducrocq y Casella, 1996). Los valores obtenidos para los parámetros de la distribución gamma en ν son relativamente altos en poblaciones de ganado lechero. Esto significa que ν aproximadamente sigue una distribución log-normal y s se encuentra distribuido aproximadamente normal (Ducrocq y otros, 1988b). Por lo tanto, Ducrocq (1997) sugirió asumir una distribución normal multivariada para s , considerando así las relaciones aditivas entre los méritos genéticos animales. Existen diversas propuestas para estimar los parámetros de la distribución de ν , siendo muy conveniente la de Ducrocq y Casella (1996), quienes sugirieron una aproximación Bayesiana para el análisis de supervivencia con modelos frágiles.

3.9. Modelos Weibull y su aplicación en el mejoramiento animal

El modelo de Cox (1972) se utilizó en el mejoramiento genético animal porque no requieren conocer la distribución de los tiempos de supervivencia, la cual es generalmente desconocida. Sin embargo, cuando un modelo de Cox tiene muchas variables que son función del tiempo y efectos fijos con un gran número de niveles, el cálculo necesario para estimar los coeficientes β es excesivamente complejo. En consecuencia, estos modelos no suelen emplearse para analizar grandes bases de datos, típicas de las evaluaciones genéticas animales. Cuando la *función de riesgo base* tiene una forma paramétrica conocida, la estimación de β y $\lambda_0(t)$ es mucho más sencilla (Cox y Oakes, 1984), por ejemplo, cuando la *función de riesgo base* posee distribución Weibull. La complejidad de estos modelos se justifica por la necesidad de describir lo más detalladamente posible, los cambios en las tendencias genéticas y los criterios de descarte a lo largo del tiempo. Por ejemplo, el efecto del cambio anual en el tamaño de un rodeo, al eliminar o incorporar vacas, responde a la necesidad de adecuarse al sistema de producción con cuotas, a las inclemencias climáticas, al manejo sanitario, u a otras causas. Desconocer estos efectos, lleva a una predicción de los valores de cría de los toros confundida con efectos ambientales.

4. EJEMPLO DE APLICACIÓN DEL ANÁLISIS DE SUPERVIVENCIA

A continuación se presentará un ejemplo de aplicación del análisis de supervivencia para predecir los valores de cría de toros lecheros, para el carácter vida productiva, utilizando los datos de Dürr, Monardes y Cue (1999).

4.1. Registros y modelo de análisis

El archivo contenía 331.147 registros de vida productiva de vacas de la raza Holstein en la provincia de Quebec, Canadá, tomados entre 1981 y 1995. El archivo de pedigree incluyó 1.875 toros de la raza Holstein. El carácter vida productiva funcional se definió como el tiempo en días desde el primer parto hasta el descarte, o al momento de la medición. Se consideró como censurados al 42% de los registros, la mayoría de los cuales correspondieron a vacas que se hallaban en producción al momento de generar la base de datos (1995). También fueron considerados registros censurados aquellos provenientes de vacas que fueron vendidas, alquiladas o exportadas a otros rodeos y los de las vacas que pertenecían a rodeos con toma de registros discontinuada.

Las predicciones de los valores de cría (VCP) se calcularon empleando un *modelo de padre*, en el que se asumió una distribución Weibull para la *función de riesgo base* (*Weibull Sire Model*). El análisis se realizó con el programa Survival Kit (Ducrocq and Sölkner, 1998b). El uso del modelo de padre en lugar del modelo animal, se debió a una limitación en la capacidad de cómputo disponible. El carácter analizado fue *vida productiva funcional*, mediante el siguiente modelo de *riesgo proporcional*:

$$\lambda(t, z) = \lambda_0(t) \exp[\mathbf{x}'(t) \mathbf{b} + \mathbf{z}' \mathbf{u}]$$

siendo $\lambda(t, z)$ el riesgo de una vaca de ser eliminada cuando ocurren los efectos considerados en los vectores $\mathbf{x}'(t)$ y \mathbf{z} ; $\lambda_0(t)$ es la *función de riesgo base*; \mathbf{b} es el vector de soluciones para los efectos ambientales en $\mathbf{x}(t)$ y \mathbf{u} es el vector de los valores de cría. Se consideraron los siguientes efectos fijos que dependen del tiempo: a) año de parición y b) cambio anual del tamaño del rodeo. Los efectos fijos independientes del tiempo fueron: c) efecto combinado del número y estado de la lactancia, d) edad al primer parto y e) método de control lechero. Los efectos aleatorios fueron: f) la mitad del valor de cría del padre de la vaca y g) rodeo-año (dependiente

del tiempo). La h^2 estimada de *vida productiva funcional* fue 0.15 (Dürr y otros, 1999).

4.2 Expresión e interpretación de los valores de cría predichos

En el ejemplo de Dürr y otros (1999), se consideró la mitad del VCP (0.5VCP), que es la fracción del mismo que un padre transmite a su progenie, expresada como *tasa de riesgo*. Por ejemplo, las hijas de un toro con $0.5VCP = 1.3$ tienen en promedio, un riesgo de ser descartadas que es 30% mayor al de las hijas de un toro con un 0.5VCP promedio ($0.5VCP = 1.0$). Alternativamente, cuando un toro tiene un $0.5VCP = 0.65$, sus hijas tienen en promedio 35% menor riesgo de ser descartadas (o 35% mayor posibilidad de sobrevivir), que las hijas de un toro promedio ($0.5VCP = 1.0$). En el presente estudio, los 0.5VCP oscilaron entre 0.52 y 1.35, siendo los valores por debajo del promedio los más deseables, dado que implican mayor supervivencia. En consecuencia, para aumentar la *vida productiva* debemos seleccionar aquellos toros que tienen los menores valores de 0.5VCP.

La variabilidad entre reproductores se puede medir indirectamente a través del cálculo de las curvas de supervivencia. Para ilustrar las diferencias genéticas entre reproductores, en la Figura 1 se grafican las curvas de supervivencia $S(t)$ para tres toros (A, B y C), con *tasas de riesgo* de descarte igual a 0.6, 1.0 y 1.3, respectivamente. Estas curvas, calculadas para una hija promedio en un rodeo promedio de la población evaluada, se pueden interpretar como la fracción de hijas de un toro que se espera alcancen cierta *vida productiva*. En la Figura 1 se puede observar que las expectativas de *vida productiva* son diferentes para las hijas de los tres toros. Por ejemplo, se espera que un 85% de las hijas del toro A estén vivas al principio de la segunda lactancia (400 días después del primer parto), cifra que para las hijas de los toros B y C corresponden al 77% y 71% respectivamente. Si se comparase al principio de la tercera lactancia (800 días después del primer parto), la expectativa de supervivencia es de 74% para las hijas del

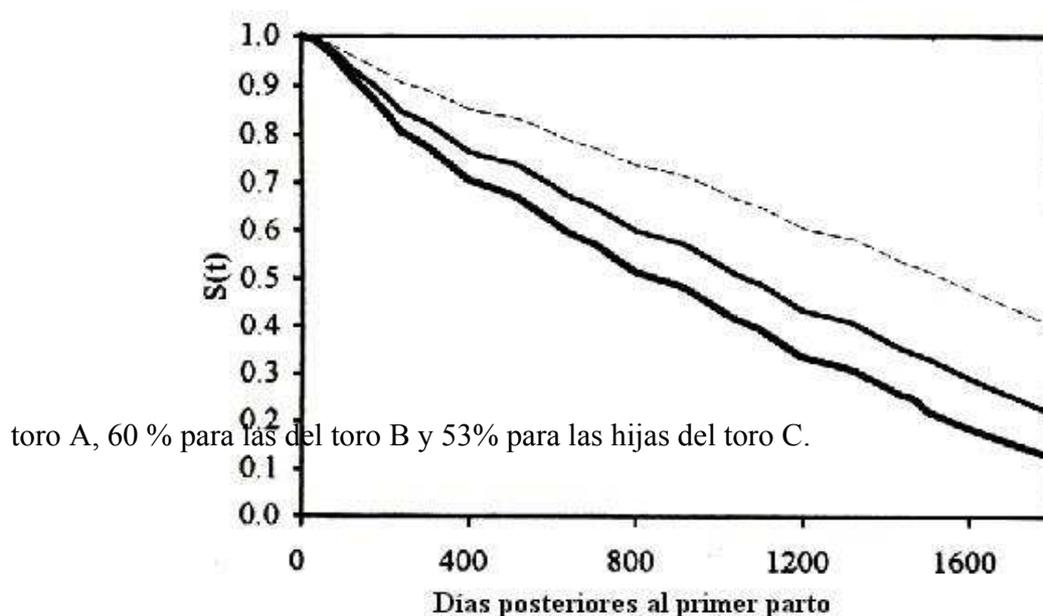
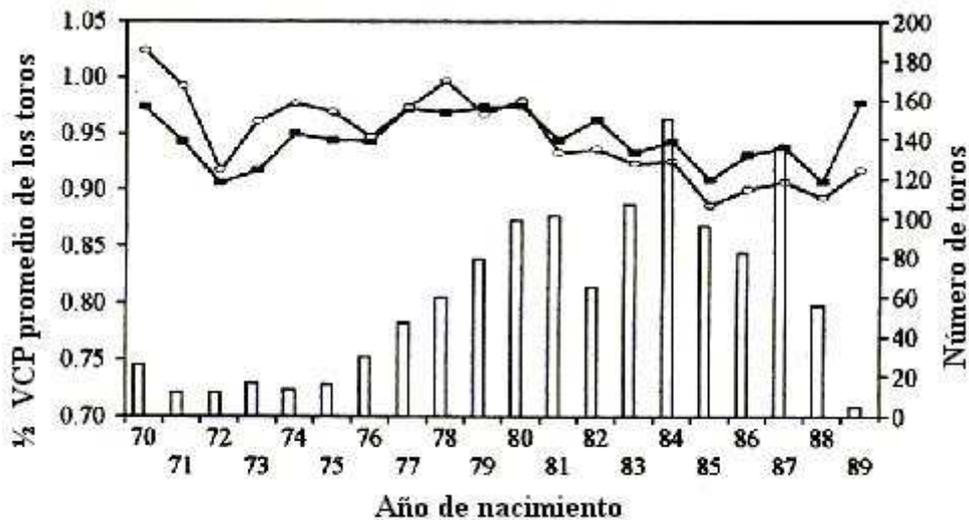


Figura 1. Curvas de supervivencia esperada para tres toros con diferentes *tasas de riesgo* de descarte. Toro A (línea a puntos) $\nu_A = 0,6$; Toro B (línea llena) $\nu_B = 1,0$ y Toro C (línea gruesa) $\nu_C = 1,3$; $S(t)$ = tiempo de supervivencia (adaptado de Dürr y otros, 1999)

Estas diferencias en supervivencia también se pueden expresar en términos del tiempo requerido para que las hijas de cada toro alcancen un valor dado de la curva de supervivencia (Cuadro 1). Por ejemplo, el tiempo medio de supervivencia de las hijas del toro C ocurre 860 días después del primer parto, mientras que el de las hijas del toro A se produce 1550 días después del primer parto, es decir, aproximadamente dos lactancias más tarde. Esta diferencia es notable e importante en relación con el impacto que tiene sobre la rentabilidad del rodeo.

Cuadro 1. Tiempo esperado (en días) al que las hijas de los reproductores A, B y C alcanzan diferentes valores de la curva de supervivencia.

Toro	$S(t) = 0.75$	$S(t) = 0.50$
------	---------------	---------------



A	150	1500
B	471	1078
C	341	860

4.3. Tendencia genética

La Figura 2 muestra la tendencia genética para los toros con más de 20 hijas con *fallas*, para *vida productiva* y *vida productiva funcional*. La tendencia genética para *vida productiva funcional* es descendente, hecho favorable porque los 0.5VCP están expresados como *tasa de riesgo* de descarte. De la Figura 2 se deduce que la selección mejoró (o al menos no deterioró) el valor de cría de los toros lecheros para los caracteres de *vida productiva* considerados. Se observaron tendencias genéticas similares para la raza Holstein en Francia (Ducrocq, 1999), en Italia (Schneider, Canavesi y Samoré, 2000) y en Alemania (Pasman y Reinhardt, 1999).

Figura 2. Tendencia genética para *vida productiva* y *vida productiva funcional*. Barras verticales = número de toros nacidos por año; rectángulos vacíos = *vida productiva* y rectángulos sólidos = *vida productiva funcional* (adaptado de Dürr y otros, 1999).

4.4. Exactitud

La exactitud de un VCP para *vida productiva* no considera el total de las hijas de un toro (con registros censurados y no censurados), sino el número de hijas en las cuales se observó el descarte (*falla*), es decir, sin datos censurados. Esto debe a la necesidad de asegurar una buena exactitud de los VCP, porque los registros no censurados corresponden a vacas que han finalizado su *vida productiva* y, en consecuencia, la información sobre el carácter es completa. La exactitud de los toros jóvenes es baja porque al momento de su evaluación poseen gran parte de sus hijas en producción, y porque la heredabilidad de *vida productiva* es baja. Los toros con todas sus hijas aún en producción tienen exactitud igual a cero, la cual puede aumentar si incorporamos la información del padre del toro (Ducrocq, 1999). El Cuadro 2 muestra diferencias de exactitud en el VCP al considerar la información de los ancestros, para toros de la raza Holstein italiana (Schneider y otros, 2000). Al incluir la información ancestral, la exactitud de los toros jóvenes aumenta entre 25 a 40%. Si consideráramos como criterio de publicación que sólo aquellos toros con una exactitud no menor al 50% sean incluidos en los resúmenes de padres, los toros jóvenes no tendrían evaluaciones. Empleando dicho criterio, los toros con mayor proporción de hijas descartadas, se incluirían antes en los resúmenes y se evaluarían en forma más precisa.

Cuadro 2. Porcentaje de toros con distintos niveles de exactitud (Fuente: Schneider y otros, 2000)

Confiabilidad	Sin información de ancestros (%)	Con información de ancestros (%)
25%	8.6	0

40%	19.6	10.4
50%	31.4	21.6
65%	60.6	52.4
75%	72.6	70.4
99%	100	100

La exactitud de los toros jóvenes se puede aumentar combinando la información de *vida productiva*, obtenida mediante el análisis de supervivencia de las hijas del toro, con la información de caracteres funcionales. Druet, Sölkner y Gengler (2000) adaptaron la metodología M.A.C.E. (Multitrait Across-Country Evaluation) utilizada por INTERBULL en las evaluaciones genéticas internacionales, para mejorar la exactitud de los VCP para *vida productiva*, combinando *vida productiva* con persistencia de la lactancia, recuento de células somáticas, fertilidad y dificultad de parto.

COMENTARIOS FINALES

El análisis de supervivencia resulta ser hasta el presente, la metodología más apropiada para evaluar reproductores para *vida productiva*, en el mejoramiento genético animal (Ducrocq, 1994, 1997; Ducrocq y Sölkner, 1998b; Kachman, 1998). Los análisis realizados empleando la metodología del análisis de supervivencia son ventajosos para analizar *vida productiva*, en relación con los modelos lineales clásicos, porque: 1) hacen uso de toda la información disponible, tanto aquella proveniente de los registros no censurados como la de los registros censurados; 2) provee un tratamiento estadísticamente informativo para los registros censurados, y 3) tienen en cuenta las características no lineales asociadas con los registros de vida productiva. Es deseable que en el futuro, las evaluaciones genéticas futuras se realicen bajo el modelo

animal, en vez del modelo de padre (sire model) o el modelo de padre y abuelo materno (maternal grand sire model). Un tópico que requiere investigación adicional, se refiere a la exactitud de los VCP de animales jóvenes. Es deseable que en el futuro, el análisis de supervivencia se pueda extender incorporando a los modelos de evaluación efectos de genes mayores o de marcadores genéticos moleculares.

BIBLIOGRAFÍA

ALLISON, P.D. 1995. Survival Analysis Using the SAS® System: A practical guide. Cary, NC.

SAS Institute Inc. 292 pp

BOETTCHER, P.J., JAIRATH, L.K. Y 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.

COX, D. R. 1972. Regression models and life tables (with discussion). J. Royal. Stat. Soc. B, 34: 187.

COX, D. R. Y OAKES, D. 1984. Analysis of survival data. Chapman and Hall, London.

DRUET, T. J. Y SÖLKNER, J., GENGLER, N. 1999. Use of multitrait evaluation procedures to improve reliability of early

prediction of survival. J. Dairy Sci., 82: 2054.

DUCROCQ, V. P. 1987. An analysis of length of productive life in dairy cattle. Ph.D. Thesis, Cornell Univ., Ithaca, NY.

DUCROCQ, V. P., QUAAS, R. L., POLLAK, E. J. Y CASELLA G. 1988a. Length of productive life of dairy cows.

1. Justification of a Weibull model. J. Dairy Sci., 71: 3061-3070.

DUCROCQ, V. P., QUAAS, R. L., POLLAK, E. J. Y CASELLA G. 1988b. Length of

productive life of dairy cows.

2. Variance component estimation and sire evaluation. *J. Dairy Sci.*, 71: 3071-3079.

DUCROCQ, V. P. 1994. Statistical analysis of length of productive life for dairy cows of the Normande breed. *J. Dairy*

Sci., 77: 855-866.

DUCROCQ, V. P. Y CASELLA, G. 1996. A Bayesian analysis of mixed survival model. *Genet. Sel. Evol.*, 28: 505-529.

DUCROCQ, V. P. 1997. Survival analysis, a statistical tool for longevity data. Book of Abstracts of the 48th Annual

Meeting of the European Association for Animal Production. Vienna, Austria.

DUCROCQ, V. P. Y SÖLKNER, J. 1998a. The Survival Kit - a Fortran package for the analysis of survival data. Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia. 27: 447-448.

DUCROCQ, V. P. Y SÖLKNER, J. 1998b. Implementation of a routine breeding value evaluation for longevity of dairy

cows using survival Analysis techniques. Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, Australia.

23: 359-362.

DUCROCQ, V. P. 1999. Two years of experience with the French genetic evaluation of dairy bulls on production-adjusted

longevity of their daughters. Proc. Int. Workshop on Genetic Improvement on Functional Traits in cattle – Longevity.

Jouy-en-Josas, France. INTERBULL Bull No 21. Int. Bull. Eval. Ser., Uppsala, Sweden, pp. 60-70.

DUCROCQ, V. P., BESBES, B. Y PROTAIS, M. 1999. Genetic improvement of laying hens viability using survival

analysis techniques. Book of Abstracts of the 50th Annual Meeting of the European Association for Animal

Production. pp. 29.

DÜRR, J. W., MONARDES, H.G. Y CUE, R.I. 1999. Genetic analysis of herd life in Quebec Holsteins using Weibull

models. J. Dairy Sci., 82: 2503-2513.

EVERETT, R. W., KEOWN, J. F. Y CLAPP, E. E. 1976. Relationships among type, production, and stayability in Holstein

cattle. J. Dairy Sci., 59: 1505-1511.

FAMULA, T. R. 1981. Exponential stayability model with censoring and covariates. J. Dairy Sci., 64: 538-545.

HARBERS, A. 1999. Durability breeding value in the Netherlands and the impact on sire rankings. Proc. of the 1999 Interbull

Meeting Zurich, Switzerland. INTERBULL Bull No 22. Int. Bull. Eval. Ser., Uppsala, Sweden, pp. 126.

JONG, G., VOLLEMA, A. R., VAN DER BEEK, S. Y HARBERS, A. 1999. Breeding values for functional longevity in the

Netherlands. Proc. Int. Workshop on Genetic Improvement on Functional Traits in cattle – Longevity. Jouy-en-Josas,

France. INTERBULL Bull No 21. Int. Bull. Eval. Ser., Uppsala, Sweden, pp. 68.

KACHMAN, S. D. 1998. Applications in survival analysis. Joint Meeting American Dairy Science Association- American

- Society of Animal Science, July 28-31, 1998. Denver, Colorado, pp147-153.
- KALBFLEISH, J. D. Y PRENTICE, R. L. 1980. The statistical analysis of failure time data. John Wiley & Sons, New York.
- KLEINBAUM, D. G. 1996. Survival analysis. A self-learning text. Springer-Verlag, New York.
- PASMAN, E. Y REINHARDT, F. 1999. Genetic evaluation for length of productive life of Holstein cattle in Germany. Proc. Int. Workshop on Genetic Improvement on Functional Traits in cattle – Longevity. Jouy-en-Josas, France. Interbull Bull No 21. Int. Bull. Eval. Ser., Uppsala, Sweden.
- RICARD, A. Y FOURNET-HANOCQ, F. 1997. Analysis of factors affecting length of competitive life of jumping horses. Genet. Sel. Evol., 29: 251-267.
- RINGMAR-CEDERBERG, E., K., JOHANSSON, K. N., LUNDEHEIM Y RYDHMER, L. 1997. Longevity of Large White and Swedish Landrace sows. Book of Abstracts of the 48th Annual Meeting of the European Association for Animal Production. Vienna, Austria. 3: 30.
- SCHAEFFER, L. R. Y BURNSIDE, E. B. 1974. Survival rates of tested daughters of sires in artificial insemination. J. Dairy Sci., 57: 1394-1400.
- SCHNEIDER, M. P. 1998. Effect of type traits on herd life in Holstein cows. M.Sc Thesis, McGill Univ., Montreal, pp 104.
- SCHNEIDER, M. P. Y MIGLIOR, F. 1999. A proposal for genetic evaluation for functional herd life in Italian Holsteins.

Book of Abstracts of the 50th Annual Meeting of the European Association for Animal Production. pp 11.

SCHNEIDER, M. P., CANAVESI, F. Y SAMORE, A.B. 2000. Genetic evaluation for functional longevity in Italian

Holsteins. Book of Abstracts of the 51th Annual Meeting of the European Association for Animal Production. pp 34.

SMITH, S. P. Y QUAAS, R. L. 1984. Productive lifespan of bull progeny groups: failure time analysis. J. Dairy

Sci. 67: 2999-3007.

STRANDBERG, E. Y ROTH, A. 1999. Genetic parameters of functional and fertility-determined length of productive life in

Swedish dairy cattle. Proc. Int. Workshop on Genetic Improvement on Functional Traits in cattle – Longevity.

Jouy-en-Josas, France. INTERBULL Bull No 21. Int. Bull. Eval. Ser., Uppsala, Sweden, pp. 152.

VANRADEN, P. M. Y KLAASKATE, E. J. H.. 1993. Genetic evaluation of length of productive life including predicted

longevity of live cows. J. Dairy Sci., 76: 2758-2764.

VOLLEMA, A. R. Y GROEN, A. F. 1998. Conformation traits in survival analysis of longevity in dairy cattle. 6th World

Congr. Genet. Appl. Livest. Prod., Armidale, Australia. 23: 371.

VUKASINOVIC, N.; MOLL, J. Y KÜNZN. 1997. Analysis of productive life in Swiss Brown cattle. J. Dairy

Sci., 80: 2372-2579.