J. Dairy Sci. 92:5730–5738
doi:10.3168/jds.2009-2132

Predictive abilities of different statistical models for analysis of survival data in dairy cattle

M. Holtsmark,*†[1] B. Heringstad,*† and J. Ødegård*‡
*Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, PO Box 5003, N-1432 Ås, Norway
†Geno Breeding and A.I. Association, PO Box 5003, N-1432 Ås, Norway
‡Nofima, PO Box 5010, N-1432 Ås, Norway

ABSTRACT

The objective of this study was to compare alternative trait definitions and statistical models for genetic evaluation of survival in dairy cattle. Data from the first 5 lactations of 808,750 first-crop daughters of 3,064 Norwegian Red sires were analyzed. Seven sire models were used for genetic analyses: linear and threshold cross-sectional models for binary survival scores from first lactation; a linear multi-trait model for survival scores from the first 3 lactations; linear and threshold repeatability models for survival scores from the first 5 lactations; a Weibull frailty model for herd life in first lactation; and a Weibull frailty model for herd life in the first 5 lactations. The models were compared to assess predictive ability of sire estimated breeding values with respect to average survival 365 d after first calving for second-crop daughters (not included in calculation of predicted transmitting abilities) of 375 elite sires. Generally, the linear multi-trait model analyzing survival in the first 3 lactations as correlated traits gave more-accurate predicted sire breeding values compared with both linear and Weibull frailty models using data from first lactation only, even when the latter models were extended to include data up to the sixth lactation. The Weibull frailty models did not improve predictive ability of sire estimated breeding values over what was obtained using a simple cross-sectional linear model for binary survival in first lactation.
Key words: predictive ability, model comparison, survival

INTRODUCTION

Longevity traits are included in the routine breeding value evaluations in many countries, and proportional hazard models or multi-trait models are most commonly used (Interbull, 2009a). Generally, a long herd life reduces replacement costs and increases the number of mature lactations. Culling can be divided into voluntary (e.g., due to low production) and involuntary culling (e.g., due to health or fertility problems). Reduced rate of involuntary culling enables farmers to increase the rate of voluntary culling such that profit is maximized, and thus may not necessarily lead to longer average herd life (van Arendonk, 1985). The economic value of reducing involuntary culling is highest in the first lactations.

True longevity has the disadvantage that it cannot be recorded before a cow is culled. This leads to right-censoring of data when breeding values are predicted while animals still are alive. In analysis of longevity traits the proportional hazards model (Ducrocq and Sölkner, 2000) has the advantage that it is able to utilize all data, including censored records. An alternative approach is to score survival as a binary trait; that is, survived or not up to a specific time, age, or event, and analyze the binary data with linear or threshold models. Use of binary survival implies that only the animals having the opportunity to survive the entire specified period can be used in genetic analysis, and records from the most recent animals will thus be excluded. Proportional hazards models analyzing time until death and treating surviving animals as censored may therefore be a better alternative. However, the most commonly used software for proportional hazards models is still restricted to univariate analyses (Ducrocq and Sölkner, 2000). In comparison, some software packages for linear or threshold models can handle multivariate genetic analysis (e.g., the DMU package; Madsen and Jensen, 2008), which enables use of information from correlated traits. Furthermore, the problems of censored lifetime data in linear or threshold models can, to some extent, be circumvented by splitting the lifespan in several intervals and define survival within each period as a binary trait (survival scores). Hence, each animal will have several records depending on the recorded lifespan, and information from younger animals can therefore be used to a greater extent. The latter method will also take into account time until death, because older animals will have several periods when they are scored

Received February 16, 2009.
Accepted August 11, 2009.
[1]Corresponding author: marte.holtsmark@umb.no

as survived before being scored as dead in the last period.

The objective was to compare alternative trait definitions and statistical models for genetic evaluation of survival in dairy cattle. Proportional hazards models (Weibull frailty) were compared with univariate linear and threshold models, using different trait definitions of survival. The models' ability to predict survival (culled or survived) in first lactation of future (second-crop) daughters of the same sires (i.e., predictive ability of sire EBV) was used for model comparison. Survival in different lactations has moderate to high genetic correlations (Jairath and Dekkers, 1994; Boettcher et al., 1999; Sewalem et al., 2007), and the aim of this study was to examine whether use of information up to the sixth lactation improves the predictive ability of the sire EBV using Weibull frailty, linear and threshold repeatability survival score models, and linear multi-trait survival score models.

MATERIALS AND METHODS

Data

Data were obtained from the Norwegian Dairy Herd Recording System. The first data set (data set 1) contained information on lactation number and calving and culling dates for the first 5 lactations of 808,750 first-crop daughters (birth-year $\leq$6 yr after the birth of the sire) of 3,064 Norwegian Red sires. Data set 1 was used to predict sire breeding values for daughter survival. The second data set (data set 2) had records on calving and culling dates for the first lactation of 1,130,255 second-crop daughters (birth year >6 yr after the birth of the sire) of 375 elite sires (sires also included in data set 1). Each elite sire had at least 1,000 second-crop daughters. Data set 2 was used for evaluation of predictive ability of the predicted sire breeding values. For both data sets, first calving had to be between 1980 and 2005, at 20 to 40 mo of age. Daughters with first calving before year 2000 were required to have a recorded date of culling, whereas the remaining cows with a missing date of culling were assumed to be alive on February 20, 2006 (truncation point). This implies that all cows had the opportunity to finish at least one lactation. In data set 1, cows with calving intervals shorter than 10 mo or longer than 24 mo, and cows that were culled more than 24 mo after last recorded calving were excluded. Cows having missing calving records before the last calving or double calving records for specific calvings were also deleted. Number of records and percentage of cows surviving per lactation are shown in Table 1. Although 68% of the cows survived the first lactation, only 5.6% of the cows survived until the end of the fifth lactation.

Table 1. Number of cows for each lactation and percentage of cows surviving the lactation

Lactation	Cows starting lactation, n	Surviving cows, %
1	800,331	68.0
2	543,898	42.7
3	341,752	23.7
4	190,062	11.7
5	93,927	5.6

In data set 1, three longevity traits were defined: 1) herd life in the first 5 lactations (**survL1–5**), calculated as the number of days from first calving to either culling from the herd or censoring (record of a sixth calving or end of recording period); 2) herd life in first lactation (**survL1**), calculated as number of days from first calving to either culling or censoring (record of a second calving or end of recording period); and 3) survival scores for lactation 1 to 5 ($\mathbf{SS_k}$, k = 1–5), with SS_k scored as 1 if the cow had a calving $k + 1$, zero if the cow was culled in lactation k, and missing if the cow was culled before the lactation or if recording period ended during lactation k. Animals culled on the same day as they calved were given a herd life of 1, and lactation survival scores were set as missing for lactations in progress at the end of the recording period. Survival scores were obtained for 799,814 first-lactation cows, which is slightly lower than the number of cows with herd life records (Table 1). Among cows with a first-lactation survival score, 536,452 survived to the second lactation, 329,381 to the third, 183,109 to the fourth, 90,262 to the fifth and 41,096 to the sixth lactation.

In data set 2, survival 365 d after first calving was scored as 0 if the cow was culled before 365 d after first calving, and 1 otherwise. In total, 26% of the second-crop daughters were scored as culled.

The sire pedigree file had records of the 3,064 sires with daughters, together with their sire and maternal grand-sires traced back as far as possible, including a total of 3,756 sires.

Weibull Frailty Model, Lactations 1 to 5

To optimize the Weibull frailty model, lactations 1 to 5 (**WFM1–5**), Kaplan-Meier curves were plotted for each of the first 5 lactations. After evaluating the curves it was decided to analyze herd life up to the sixth lactation (survL1–5) using the following sire model:

$$h_{ijklmn}(t) = h_0(t) \exp \left\{ \begin{matrix} AFC_j + YM_k + LACTSTAGE(t_1) \\ + hy_i(t_2) + sire_m \end{matrix} \right\}, [1]$$