J. Dairy Sci. 98:1–14
http://dx.doi.org/10.3168/jds.2014-9090

Genetic analysis of longevity in Dutch dairy cattle using random regression

M. L. van Pelt,*†[1] T. H. E. Meuwissen,‡ G. de Jong,* and R. F. Veerkamp†
*CRV BV, Animal Evaluation Unit, PO Box 454, 6800 AL Arnhem, the Netherlands
†Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 338, 6700 AH Wageningen, the Netherlands
‡Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, PO Box 5003, N-1432 Ås, Norway

ABSTRACT

Longevity, productive life, or lifespan of dairy cattle is an important trait for dairy farmers, and it is defined as the time from first calving to the last test date for milk production. Methods for genetic evaluations need to account for censored data; that is, records from cows that are still alive. The aim of this study was to investigate whether these methods also need to take account of survival being genetically a different trait across the entire lifespan of a cow. The data set comprised 112,000 cows with a total of 3,964,449 observations for survival per month from first calving until 72 mo in productive life. A random regression model with second-order Legendre polynomials was fitted for the additive genetic effect. Alternative parameterizations were (1) different trait definitions for the length of time interval for survival after first calving (1, 3, 6, and 12 mo); (2) linear or threshold model; and (3) differing the order of the Legendre polynomial. The partial derivatives of a profit function were used to transform variance components on the survival scale to those for lifespan. Survival rates were higher in early life than later in life (99 vs. 95%). When survival was defined over 12-mo intervals survival curves were smooth compared with curves when 1-, 3-, or 6-mo intervals were used. Heritabilities in each interval were very low and ranged from 0.002 to 0.031, but the heritability for lifespan over the entire period of 72 mo after first calving ranged from 0.115 to 0.149. Genetic correlations between time intervals ranged from 0.25 to 1.00. Genetic parameters and breeding values for the genetic effect were more sensitive to the trait definition than to whether a linear or threshold model was used or to the order of Legendre polynomial used. Cumulative survival up to the first 6 mo predicted lifespan with an accuracy of only 0.79 to 0.85; that is, reliability of breeding value with many daughters in the first 6 mo can be, at most, 0.62 to 0.72, and changes of breeding values are still expected when daughters are getting older. Therefore, an improved model for genetic evaluation should treat survival as different traits during the lifespan by splitting lifespan in time intervals of 6 mo or less to avoid overestimated reliabilities and changes in breeding values when daughters are getting older.

Key words: dairy cattle, survival, longevity, random regression

INTRODUCTION

Longevity is of economic importance for dairy farmers because increased longevity helps to increase profitability by (1) reducing the annual costs of replacement of cows; (2) increasing the average herd yield through an increase in the proportion of cows in the higher-producing age groups; (3) better use of a given acreage by reducing the number of replacements to be reared, and therefore allowing an increase in size of the milking herd; and (4) increasing voluntary culling (Rendel and Robertson, 1950). Van Arendonk (1985) showed that when involuntary culling decreases, a higher voluntary culling rate can be applied, resulting in a larger profit for the farmer but not necessarily increasing lifespan as a whole. Hence, an important economic aim for the farmer should be to reduce involuntary culling by using genetic selection.

In animal breeding, selection for longevity is complex because the true longevity of a cow is available only at the end of her life, whereas selection and breeding decisions are made earlier in life. Therefore, censored data are used in genetic evaluations, and expected lifespan needs to be extrapolated from survival data during life. To enable this analysis of censored data, different genetic evaluation models are used (Forabosco et al., 2009). Survival analysis with a proportional hazard function is a popular model used in genetic evaluations for longevity, since the software package Survival Kit was introduced in 1994 (Ducrocq and Sölkner, 1994, 1998; Ducrocq et al., 2010). That model attempts to estimate the probability that an animal will survive to time t given that it has survived to time $t - 1$.

Received November 10, 2014.
Accepted March 8, 2015.
[1] Corresponding author: mathijs.vanpelt@wur.nl

Although current genetic evaluations are restricted to a single genetic effect during life, in the newest release it is possible to analyze 2 correlated random effects (Mészáros et al., 2013).

An alternative approach to the survival model is to score survival as a binary trait; for example, survived or not up to a specific endpoint (i.e., time, age, lactation) or survived or not in a specific interval of time. This binary trait for survival can then be analyzed with a linear model or a threshold model, although correlations between breeding values based on a linear and a threshold model are very high and no or little reranking is expected (Boettcher et al., 1999; Holtsmark et al., 2009), which has also been shown for other traits (Meijering and Gianola, 1985; Carlén et al., 2006). Therefore, several countries apply a linear model with survival defined as a binary trait in their routine national genetic evaluations (Forabosco et al., 2009).

If survival is scored as a single binary trait, however, information from culling before the endpoint or information from survival after the endpoint is ignored. In addition, information from living animals that did not have the opportunity to survive a certain specified time interval cannot be used in the analyses. To overcome these major limitations, an expansion of censored records to expected lifespan has been suggested and implemented in the UK (Brotherstone et al., 1997), and data augmentation (Guo et al., 2001) or multi-trait models including survival to several endpoints have been suggested (Boettcher et al., 1999; Sewalem et al., 2007). The advantage of the multi-trait model is that information for survival after first calving of younger animals still alive can be used to score the traits of the first intervals (they already survived), and the traits for the later intervals can be set to missing; also, variation across time is included in the genetic evaluation. Ideally, many traits with different endpoints should be included in the analysis, but when the life of a cow is split in many binary traits, the multi-trait parameterization becomes impossible and random regressions models might be useful to find a more parsimonious parameterization (Schaeffer and Dekkers, 1994). For this reason, Veerkamp et al. (2001) proposed the random regression model (**RRM**) for survival analysis, which was further investigated by Jamrozik et al. (2008). Veerkamp et al. (2001) proved, by comparison of the likelihood, the equivalence between the survival model and the RRM when a large number of survival intervals were defined in the random regression model. Compared with the survival model, the advantage of the RRM is that, in a linear model context, multiple genetic effects can be fitted. Random regression models therefore enable the modeling of longevity as a different but genetically correlated trait across the entire productive life of a cow.

The assumption that longevity is genetically the same trait during the entire productive life of a cow is often violated, because evidence shows that survival is not necessarily the same trait during the total lifespan of a cow, phenotypically as well as genetically. Cows in higher parities have a higher culling risk compared with earlier parities (Ducrocq, 2005; Terawaki and Ducrocq, 2009; De Vries et al., 2010) and dairy producers cull more intensively for low production in first than in later parities (Dürr et al., 1997, 1999). Studies using linear multi-trait models showed genetic correlations between parities that significantly differ from unity, averaging 0.77, with a range from 0.33 to 0.96, where correlations decrease with increasing distance between parities (Visscher and Goddard, 1995; Brotherstone et al., 1997; Boettcher et al., 1999; Veerkamp et al., 2001; Sewalem et al., 2007; Holtsmark et al., 2009). Similar results were found with a lactation RRM (Gengler et al., 2005). Hence, this supports the question of whether the use of the survival model with 1 (or 2) genetic effects over the entire life of a cow is optimal, and whether a linear RRM with many time intervals across the life of a cow provides an alternative. A first step to adopting a new model for genetic evaluation is to determine if survival is genetically the same trait across the entire life of a cow.

Therefore, the objective of this study was to estimate genetic parameters for survival across the life of Dutch dairy cows using random regression and multi-trait models of different orders. However, a concern might be that when the lifespan of a cow is split into many short time intervals, the risk of culling becomes too low in each time interval to apply a linear model. Therefore, both a generalized linear model with a logit link function (GLMM) and an ordinary linear model ignoring the binary nature of the trait were used, and 4 different lengths of time interval for survival were evaluated.

MATERIALS AND METHODS

Data

Productive life or lifespan was defined as the time from first calving to the last test date for milk production before the animal died or was culled for slaughter (including dry periods). Data were available from the Dutch/Flemish cattle improvement cooperative CRV (CRV, Arnhem, the Netherlands). The data set was constructed from records of pedigree, lactations, and movements of cows in the Netherlands. Records were constructed for each month that a cow had been present in a herd since the first calving. A cow, culled in month j has $j - 1$ records with score 100 (alive) and record j with score 0 (culled). After culling, monthly