Longevity and Fertility Trait Definitions Compared in Theory and Simulation

P.M. VanRaden

Animal Improvement Programs Laboratory, Agricultural Research Service,
United States Department of Agriculture, Beltsville, MD, USA 20705-2350
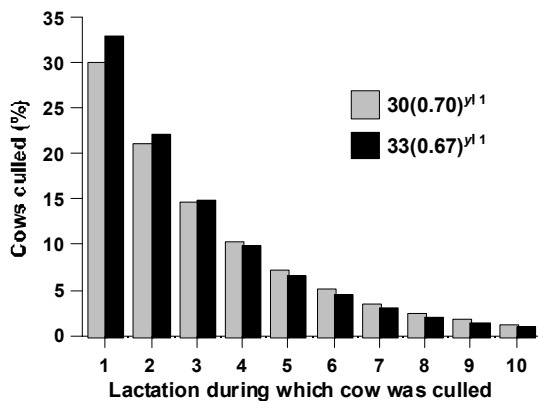
Abstract

Longevity and fertility traits often are based on binomial observations that are repeated only if the cow is not yet pregnant or culled. Time until pregnant or time until culled are not normally distributed, but theory indicates that linear models can extract all information from such data. With binomial models, cows with more observations are assumed to provide more information about the success rate parameter. Although reported reliabilities differ slightly for binomial models as compared with productive life models, estimated breeding values were correlated by 0.991and had identical true accuracies as obtained from simulated data. For traits with low heritability, information from progeny-tested maternal brothers can lead to larger benefits from animal models as compared to sire models.

Keywords: Longevity; Binomial traits; Genetic evaluation; Dairy cattle

Trait distributions

Longevity and fertility traits measure binomial events that occur across time. At any particular time, cows are either pregnant or not pregnant, alive or dead. Such traits can be defined as time until pregnant and length of life, rates of conception and survival, or pregnancy and survival status at a given time or at several points in time. Whether binomial or continuous, none of those definitions provide normally distributed data.

For some traits with very non-normal distributions, such as number of lactations until culling (Figure 1), simple linear functions of the data may still provide excellent parameter estimates. Longevity and fertility traits often are distributed as functions of success rate parameters rather than mean and variance parameters of normal distributions. For example, data on length of life from n cows may be modeled as a function of culling rate. If parameter c = fraction of cows culled in each lactation and y_i = number of lactations survived by cow i, the distribution of observation y_i is

$$\text{Prob}(y_i) = c(1-c)^{y_i-1}.$$

The joint probability for data from n cows is the product of n individual probabilities. The maximum likelihood (ML) estimate of parameter c ($\hat{c}$) can be obtained by setting the derivative of the natural logarithm (log) of the joint density with respect to c equal to 0:

$$n/\hat{c} = [(\Sigma y_i)\text{-}n]/(1\text{-} \hat{c}),$$

Figure 1. Theoretical distribution of longevity as measured by number of lactations until culling (y) at rates of 30% or 33%.

which then gives:

$$\hat{c} = n/\Sigma y_i.$$

The ML estimate $\hat{c}$ is not a linear function of the data because Σy_i appears in the denominator rather than in the numerator. However, the ML estimate of c can be obtained as a simple non-linear function of the parameter estimate for the number of lactations (t):

$$\hat{t} = \Sigma y_i/n.$$

That linear model estimate for number of lactations can be converted to a non-linear estimate of culling rate simply by taking the reciprocal:

$$\hat{c} = 1/\hat{t}.$$

A linear model analysis of longevity followed by non-linear conversion to survival rate after analysis provides estimates equivalent to those from the non-linear model in this simple case. Both procedures provide the ML estimate of c with no loss of information even though the longevity data are not normally distributed.

Longevity and fertility traits may be defined as continuous or discrete variables. To account for partial lactations, y_i can be assigned real rather than whole integer values.

Prior distributions

True breeding values (BV) expressed on different scales may have product-moment correlations of <1 because of non-linear relationships of parameters even if sires rank exactly the same on all scales. Absolute values of correlations were above 0.98 for BV expressed as number of lactations survived, culling rate, or log of relative culling rate. Lubbers et al. (2000) reported correlations of 0.96 to 0.97 between estimated breeding values (EBV) for number of lactations from a linear model and Weibull model EBV from survival analysis.

If BV are normally distributed on a given scale, they are non-normally distributed on other scales unless the transformation between scales is linear. Normally distributed BV are not quite correct in theory either for number of lactations, or culling rates, or log of culling rate because very small positive probabilities are assigned to culling rates that are negative or >1. However, that same problem occurs with traditional normal priors for milk yield, which assume that negative yields may be possible.

Differences from normality are small when heritabilities are low. If the prior distribution is nearly normal for log of relative risk, it becomes slightly skewed when translated to the other scales. The distribution has an extended right tail when expressed as culling rate and an extended left tail when expressed as number of lactations. The net result is slightly different shrinkage of daughter means toward the population mean when the daughter means are above or below population mean.

Reliability

Genetic effects are estimated more accurately from repeated records than from single records. For example, reliability of milk yield evaluations increases when later lactation records are added. Similarly, a cow that survives many years may contribute more information to her own and to her sire's longevity evaluation than a cow that is culled after just one lactation. Reliability of fertility may be higher for cows that require several services to conceive than for those that become pregnant on the first insemination. Those facts hold true even when all records are complete (no censoring).

Analyses of repeated binomial data, such as the longevity evaluations of Canada, Ireland, and Australia, have higher reliability for cows that lived longer because such cows have more observations. Thus, prediction error variances (PEV) are slightly smaller for sires with high EBV than for those with low EBV. Analysis of binomial data can provide different estimates of reliability even though EBV may be nearly identical to EBV from number of lactations. That hypothesis was confirmed with simulated data in the next section.

Heritability from binomial analyses often is expressed per lactation and thus is lower and cannot be compared directly with lifetime heritabilities reported from most other analyses. To obtain comparable reliabilities from multiple-trait, across-country evaluation, countries with repeated binomial models could either convert heritabilities

to a lifetime basis, or report effective daughter contributions of >1 for daughters with multiple binomial records.

Simulation

Data were simulated with a simple sire model similar to that of Yazdi et al. (2002). Mean culling rate was 33% per lactation, and no fixed effects were included. Sire effects were normal for log of (culling rate/0.33) and had a standard deviation of 0.28 for true BV. Variance components and sire EBV were estimated using PROC MIXED of SAS (1999). Each replicate included 76,919 completed longevity observations from 500 sires with 10 to 1000 daughters each. Median number of daughters was 26, and mean number of daughters was 154. Means and standard deviations were obtained for 10 replicates.

Transformation of data to log of number of lactations resulted in slightly lower heritability (0.091 versus 0.106) and lower accuracy (0.70 versus 0.72) as compared with EBV on the original scale. Lubbers et al. (2000) also estimated a lower heritability with log transformation of actual data (0.05 versus 0.06). Correlation of EBV on the two scales was 0.982 from simulated data and 0.97 from the actual data of Lubbers et al. (2000).

Log transformation can be helpful to increase normality and heritability for many right-skewed distributions such as somatic cell count. For longevity, transformation was not helpful because each additional year of life provides more information about the culling rate parameter. An analogy is determining the probability that a tossed coin will land with the head side up. The numbers of tosses required to obtain a head can simply be averaged without taking the log of number of tosses.

Heritability of binomial culling rate per lactation was 0.038 as compared with a heritability of 0.106 for number of lactations. The EBV for number of lactations and EBV for binomial culling rate from the same data were correlated by 0.991 and had identical actual accuracy (0.72). Thus, the larger number of binomial observations (three times as many) compensates for the lower heritability per observation. Similar reduced heritabilities for

productive life predictions at early ages were obtained by VanRaden and Klaaskate (1993). Their estimated heritabilities were 0.085 at 7 years of age but only 0.031 at 3 years because data were nearly binomial at that age.

Separate genetic effects across time can be estimated in repeated binomial models. Such models for longevity are analogous to test-day models for yield traits. Most linear analyses of productive life include only a single BV, although the New Zealand evaluation estimates multiple BV across time. Survival model evaluations include only a single EBV for each sire, but the effect is scaled by changing culling rates across time.

Survival analyses account for changing culling patterns, but a potential problem is that cows with poor fertility may be credited with greater longevity. Such cows spend a greater fraction of their lives in the late-lactation group, for which culling rates are high. Thus, they may be credited with resisting more culling than cows with short calving intervals. That problem is complex and was not investigated further because the exact adjustments included in survival analyses might differ from those simulated.

Maternal relatives

Compared with sire-maternal grandsire models, animal models may have a greater advantage when heritabilities are low. For a trait with heritability of 4%, parent average can be worth up to 100 progeny records. Proven bulls in the United States now have a mean of 5 maternal brothers that also are proven. The extra information provided by maternal brothers and sisters and the dam is worth an extra 20 daughter records for fertility as compared with including only the maternal grandsire of a bull.

Boettcher et al. (1999) found larger differences between animal and sire models than between linear and non-linear models. To offset lost information from an incomplete relationship matrix that includes only sires or only sires and maternal grandsires, the added precision from non-linear models needs to be large. More countries now use linear animal models than non-linear sire models (Van der Linde and de Jong, 2003).

Conclusions

Longevity and fertility traits can be expressed and evaluated in several ways. Simulation indicated that repeated binomial evaluations can provide EBV nearly identical to EBV for number of lactations, but that reported heritabilities and reliabilities differ somewhat. The heritability per lactation is lower in repeated binomial models than in productive life models, but the number of observations is greater, providing equivalent total information. In binomial analyses, animals that live longer have higher reported reliabilities because their repeated records are assumed to provide more information about the culling rate parameter. In productive life analyses, reported reliabilities are means instead of being slightly lower or higher for animals that had shorter or longer lives.

Estimates of culling rates with ML properties can be obtained using linear models even though data are not normally distributed. For example, the probability that a cow will be culled during a given lactation equals the per-lactation culling rate multiplied by the percentage of cows that survived to begin that lactation. Although the distribution of longevity measured as number of lactations is right skewed, transformation to reduce skewness is not helpful because long survival times provide more information. In this case, Bayesian, binomial, and standard linear estimates all depend on the data only through the mean number of lactations, which is a linear function of the data.

Prior distributions for these different analyses differ slightly. True BV for number of lactations, culling rate, and log of culling rate were highly correlated when a heritability of 0.10 was assumed. Accuracy of EBV was similar when simulated data were evaluated as binomial observations or as number of lactations. Larger differences are expected from actual data because of differences in modeling fixed effects across time or adjusting for different correlated traits. When evaluating longevity and fertility traits with low heritabilities, precise modeling may not be as important as including all relatives in an animal model.

References

Boettcher, P.J., Jairath, L.K. & Dekkers, J.C.M. 1999. Comparison of methods for genetic evaluation of sires for survival in the first three lactations. *J. Dairy Sci. 82,* 1034-1044.

Lubbers, R., Brotherstone, S., Ducrocq, V.P. & Visscher, P.M. 2000. A comparison of a linear and proportional hazards approach to analyze discrete longevity data in dairy cows. *Anim. Sci. 70,* 197-206.

SAS OnlineDoc®, Version 8. 1999. SAS Inst., Cary, NC.

Van der Linde, C. & de Jong, G. 2003. MACE for longevity traits. Proc. Interbull Tech. Workshop, Beltsville, MD, Mar. 2-3. *Interbull Bulletin 30,* 3-9.

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.

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.

Acknowledgments