

IMPACT ON CALVING EASE EVALUATIONS OF MODIFYING OR EXCLUDING DATA FROM HERD-YEARS WITH ABNORMAL DISTRIBUTIONS OF SCORES

C.P. Van Tassell^{1,2} and G.R. Wiggins¹

¹Animal Improvement Programs Laboratory and ²Bovine Functional Genomics Laboratory
Agricultural Research Service, USDA Beltsville, MD, USA.

INTRODUCTION

A sire-maternal grandsire threshold model for evaluation of calving ease was introduced in the United States in August 2002 (Van Tassell *et al.*, 2003). This evaluation is based on owner assigned scores of 1 to 5, with 1 being unobserved or no difficulty, and 5 being extreme difficulty. Investigation of evaluations that had large changes from run to run revealed that data for these bulls usually was from herds with abnormal distributions of scores. Some owners have made 3 the most common score assuming that the scores should be centrally distributed. In other cases, most or all of the observations were 5, the most extreme value. Threshold model evaluations are sensitive to extreme values. These issues related to the distribution of scores suggested that the evaluations might be more stable if herd-years with abnormal distributions were excluded or scores were transformed so their distribution was more typical.

When proposing changes to an evaluation system where the true breeding values are unknown, a means of assessing the benefit of a change is needed. Prediction of future evaluations is a useful measure. A modification of the data is expected yield evaluations that increase the correlations between evaluations calculated from the complete data and those from a subset having recent records removed. The objective of this study was to determine if some combination of reassigning score categories to force a mode of one and removing records from herd-years with abnormal distributions increased the consistency of evaluations over time.

MATERIAL AND METHODS

The 13,303,751 calving ease records included in the August 2005 evaluation were the basis for this study. For each of three time periods, evaluations were calculated from six data sets: all data, data transformed to force the modal score to be one, and data further reduced by two levels of exclusion for each of two goodness-of-fit (GoF) tests. The time periods included data through July 2005 calvings (all), through September 2004 (previous) and through May 2001 (truncated). These periods were chosen to represent all available data, data included in the evaluation six months earlier (the frequency of calculation of national genetic evaluations), and data with the last four years excluded to enable determination of the value of ancestor information in predicting current evaluations where no direct information is available for the bull.

For each herd-year, the modal score was determined. Where the mode was two, scores two and greater were reduced by one. This combined the original scores of one and two. Similarly for a mode of three, the first three scores were combined and the remaining scores were reduced by two. For modes of four or five, the data for the herd-year were excluded. These adjustments to observed data were based on the premise that scorers of dystocia were centralizing the data because they did not fully understand the scoring scheme, but their scores did provide reliable rankings. The number of calvings in herd-years where adjustments were applied was 1,118,052. These records represent 8.8% and 6.3% of herd-years for parity 1 and 2+,

respectively. A total of 157,879 records were removed because they were in herd-years with modes of 4 or 5, representing 1.6% and 1.0% of parity 1 and 2+ herd-years.

Two sets of goodness-of fit (GoF) statistics were calculated for each herd-year group. The first GoF statistic calculated, GoF1, is based on the logarithm of the likelihood ratio of the observed distribution of observations compared to the likelihood of the same number of observations distributed with population frequencies by parity. The likelihood assumes a multinomial distribution. Specifically, the values were calculated as:

$$GoF_1 = \sum_{j=1}^2 \sum_{i=1}^5 (n_{i,j} - n_{\cdot,j} f_{i,j}) \ln(f_{i,j}) - \ln(\Gamma(n_{i,j} + 1)) + \ln(\Gamma(n_{\cdot,j} f_{i,j} + 1)),$$

where $f_{i,j}$ is the fraction of calving ease score i in parity j across the entire population (see Table 1), $n_{i,j}$ is the number of score i parity j calvings for a herd-year group.

The second GoF statistic is based on the Kolmogorov-Smirnov non-parametric test (Boes *et al.*, 1974) for goodness of fit. First, the empirical distribution function is defined for a set of n independent observations X_1, X_2, \dots, X_n from cumulative distribution function $F(x)$, order the observed samples from lowest to highest and let $X_{(i)}$ be observation i of the sorted values. Then, define

$$F^*(x) = \begin{cases} 0, & x < X_{(1)} \\ \frac{i}{n}, & X_{(i)} \leq x < X_{(i+1)} \quad i = 1, \dots, n-1 \\ 1, & x \geq X_{(n)} \end{cases}$$

Finally, the Kolmogorov-Smirnov statistic (D) is defined as

$$D = \sup_x |F^*(x) - F(x)|.$$

For application to the calving ease data edits, a shortcoming of this statistic was noted; the ordered nature of the data is not considered, just the largest discrepancy between the population and herd-year frequencies. For example, consider the sample data in Table 1, where one change is made from parity 1 frequencies in score 1 from population values and that fraction was moved to scores 2 or 5. These two scenarios reflect dramatically different potential impacts on genetic evaluations, however, the Kolmogorov-Smirnov D (**KSD**) statistic is identical for those two examples.

The KSD values were modified to reflect the “distance” that the mass of probability moved – so for the two sample herd-years, Sample 2 test statistic is four-times the value for Sample 1.

Table 1. Distribution of calving ease scores by parity group and example first parity data

Score	Population Frequency		Example First Parity Data	
	Parity 1	Parity 2	Sample 1	Sample 2
1	63.9	82.7	33.9	33.9
2	14.9	8.9	44.9	14.9
3	13.8	5.8	13.8	13.8
4	4.9	1.7	4.9	4.9
5	2.5	0.9	2.5	32.5

Herd-years were divided into 100 groups of increasing herd size that contained nearly equal numbers of calving records. Thresholds for 1, 5, and 10 percent were calculated for each group. Fitting equations which changed at herd sizes 141 and 268 smoothed these values. These smoothed values then were used to determine which herd-years to exclude.

To avoid the undesirable situation where a herd's data often switched from being included to being excluded in consecutive years, the 10 percent threshold was used to exclude herd-years adjacent to an excluded herd-year. This additional exclusion was applied both forward and backward in time, so one excluded herd-year could cause many years to be excluded for a herd. The 10 percent level was used with both the 1 and 5 percent exclusion levels test resulting in actual exclusions of 3 and 8 percent of data.

The sire-maternal grandsire model used for the routine genetic evaluation (Cole *et al.*, 2005) was used; the same pedigree data was used for all evaluations. This caused evaluations to be calculated for bulls that did not have any progeny births or daughter calvings. These evaluations were based on ancestors, so were appropriate for determining the effectiveness of the evaluation in predicting future evaluations.

Evaluations were compared by calculating correlations between the pre-May 2001 evaluations and those using all years for the 3 levels of exclusion. For service sire calving ease, Holstein AI bulls with reliability in the complete evaluation of 70 or more were used. To include approximately the same number of bulls for daughter calving ease, the reliability requirement was reduced to 60.

RESULTS AND DISCUSSION

Several response variables were evaluated in characterizing response to data edits, including: correlations, mean squared error, average absolute difference, maximum change, a number of percentile changes, and standard deviation of changes. A number of measures of stratifications were also used in an attempt to visualize results, including change in numbers of new records, fraction of new records, and change in reliability.

The simple adjustment to modes had noticeable impact on improving evaluation stability when comparing current and truncated genetic predictions. Beyond that, results showed different methods are best, depending on the criteria used for comparing data edits. Particularly surprising were the differences when comparing results over near and longer intervals for similar changes in reliability. The increased variability of correlations at the highest changes in reliability reflects the small numbers of bulls represented.

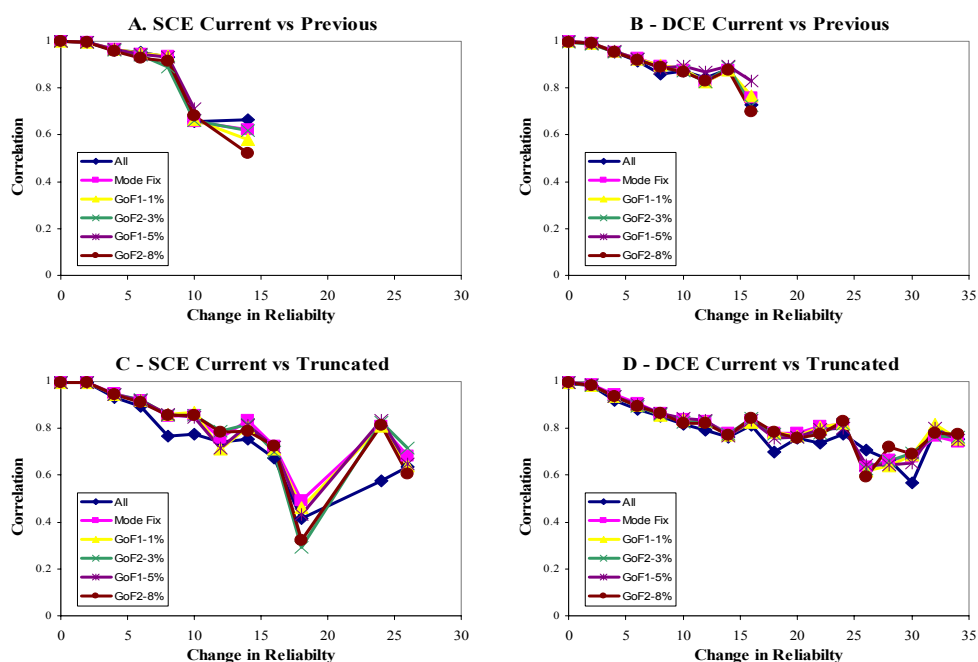


Figure 1. Correlations between sire calving ease (SCE) and daughter calving ease (DCE) for current vs truncated and previous evaluations for Goodness of Fit (GoF) 1 and 2 for varying levels of data exclusion.

CONCLUSION

These results show, depending on the criteria used for comparing data edits, different methods are best. Particularly surprising were the differences when comparing results over near and longer intervals for similar changes in reliability. Work continues on identifying GoF measures and means to measure response to these edits.

REFERENCES

- Boes, D.C., Graybill, F.A. and Mood, A.M. (1974) << Introduction to the Theory of Statistics >>. 3rd ed. McGraw-Hill New York USA.
- Cole, J.B., Goodling, R.C., Wiggans, G.R. and VanRaden, P.M. (2005) *J. Dairy Sci.* **88** : 1529–1539.
- Van Tassell, C.P., Wiggans, G.R., and Misztal, I. (2003) *J. Dairy Sci.* **86** : 3366–3373.