

## EFFECT OF DATA QUALITY ON GENETIC PARAMETERS FOR DAYS TO FIRST INSEMINATION IN DAIRY CATTLE

G. E. Shook<sup>1</sup>, K. A. Weigel<sup>1</sup>, J. Sun<sup>1</sup> and J. S. Clay<sup>2</sup>

<sup>1</sup>Dairy Science Department, University of Wisconsin, 1675 Observatory Drive, Madison, Wisconsin 53706 USA

<sup>2</sup>Dairy Records Management Systems, Raleigh, North Carolina, 27695 USA

### INTRODUCTION

Interval from calving to first breeding is regarded as one of the most important practical measures of reproductive performance (Thaller, 1998). Extended intervals may be caused by post-partum metritis, endocrine dysfunction that delays the establishment of regular estrus cycles, and depressed expression of estrus behavior. Undetected estrus is one of the major contributors to extended intervals to first breeding or conception.

Reproductive event data submitted by commercial farms is quite variable in its completeness and accuracy. Data quality at the herd level has been measured by the percentage of records with normal gestation length (GL), defined as  $259 \leq GL \leq 301$ , and by the standard deviation of GL (Zhang and Shook, 2000 ; Zhang and Shook, 2001). Gestation length was chosen as a measure of data quality because it depends on accurate reporting of insemination and calving dates, is not substantially influenced by management, and has a low coefficient of variation.

An optimal level of data quality has not been established. When the quality standard is low, the value of individual cow records is diminished, but the value of aggregated records, as for sires, may be acceptable due to the large number of animals evaluated. Conversely, when quality standards are high, the value of individual cow records is improved, but the value of aggregated records could be decreased due to the small number of animals evaluated. An intermediate quality standard may lead to the highest rate of genetic progress.

The frequency distribution of days to first breeding has positive skewness (i. e., the mean is greater than the median). The objectives of this study were to determine the effects of alternative data quality standards and data transformation on the genetic parameters of days to first breeding (DFB).

### MATERIALS AND METHODS

Data from herds on Dairy Herd Improvement throughout the U. S. were compiled and provided by Dairy Records Management Systems, Raleigh, NC. Records included in the analysis were from parities 1 to 4, herds with first service non-return rate between 10 and 90%, and herds with  $\geq 20$  records. Only records with  $30 \leq DFB \leq 240$ , identified sire and identified mating sire were retained. After editing, 4,531,536 records were available for analysis.

From a preliminary analysis, we chose  $LDFB = \text{Log}_2 [DFB/15]$  as the transformation for DFB to obtain a more nearly Normal frequency distribution. The inverse transformation is  $DFB = 15 * 2^{LDFB}$ . The LDFB scores of 1, 2, 3 and 4 correspond to DFB 30, 60, 120 and 240.

Genetic parameters were estimated with the REMLF90 software (Misztal, 2000) using the mixed model:  $Y = Xb + Zu + Wp + Ie$

Where:

Y is a vector of observed values, DFB or LDFB;

b is a vector of fixed effects, herd-year-season and age-parity;

u is a vector of random animal effects, breeding values;

p is a vector of random permanent environmental effects;

e is a vector of random residuals;

X, Z and W are incidence matrices that link b, u and p to the observed values, Y;

I is the identity matrix.

Variances of the random effects were  $A_u \sigma^2(u)$  for animals,  $I\sigma^2(p)$  for permanent environment and  $I\sigma^2(e)$  for residuals, where  $A_u$  is the genetic relationship matrix among animals and their close relatives, including males with progeny.

The within herd percentage of records with normal GL [NGL%] and standard deviation of GL [SDGL] were computed for each herd. Quintiles were determined for each measure. Quintile boundaries were 61, 68, 74 and 79% for NGL% and 5.5, 5.9, 6.4 and 6.9 days for SDGL. Genetic parameters were estimated for 9 of the 25 quintile combinations of NGL% and SDGL. Only combinations that involved the first, third, and fifth quintiles of both measures were analyzed. The product-moment correlation between NGL% and SDGL was  $-0.24$ .

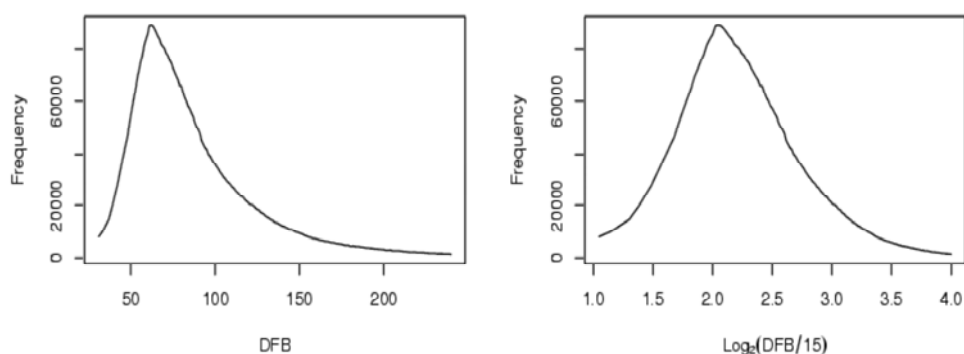
## RESULTS AND DISCUSSION

Descriptive statistics for DFB and LDFB are in Table 1. Under the Normal frequency distribution, expectations for skewness and kurtosis are 0 and 3.0. Skewness and kurtosis for LDFB were substantially closer than those for DFB to the Normal expectations. Frequency distributions for DFB and LDFB are shown in Figure 1.

**Table 1. Descriptive statistics for days to first breeding (DFB) and the log transformation of days to first breeding (LDFB)**

	Mean	Std. Dev	Skewness	Kurtosis
DFB	85.2	35.9	1.41	5.16
LDFB	2.39	0.552	0.373	2.88

The numbers of observations in the nine data sets are shown in Table 2. There were an average of 8.8 records per contemporary group, 139 animals with records per herd and 221 records per herd in these data sets.



**Figure 1. Frequency distributions for days to first breeding (DFB) and the log transformation of days to first breeding (LDFB)**

**Table 2. Numbers of animals, records, and herds for selected combinations of NGL% and SDGL quintiles**

Quintile		Numbers of		
NGL%	SDGL	Animals <sup>A</sup>	Records	Herds
1	1	38,508	65,242	319
1	3	51,011	84,699	262
1	5	69,034	109,163	406
3	1	60,696	99,030	519
3	3	91,363	145,700	432
3	5	35,992	54,789	283
5	1	143,882	228,563	1,237
5	3	67,542	103,926	455
5	5	18,361	26,438	233

<sup>A</sup>Animals with records

The estimated genetic parameters for DFB are in Table 3. Variances for animal, permanent environment and residual declined as NGL% increased. There was no consistent trend for the variances with respect to SDGL. Over the 9 data sets, heritability averaged 4.5% and repeatability averaged 10.0% with little variation among data sets. No important trend was seen in heritability or repeatability estimates with respect to the quintiles for NGL% and SDGL.

Heritability and repeatability for LDFB differed slightly from those for DFB [data not shown]. In individual data sets, these variance ratios differed by less than 1%. Heritability estimates in these data are similar to other studies; typical estimates have been 3%, 3%, 6% and 6% (Hayes et al., 1992 ; Pryce et al., 1997 ; Pryce et al., 2001 ; Weigel and Rekaya, 2000).

**Table 3. Estimated variances, heritabilities ( $h^2$ ), and repeatabilities ( $r$ ) from mixed model analysis for days to first breeding by quintile groups for NGL% and SDGL**

Quintiles		Variances				
NGL %	SDGL	Animal	PE <sup>A</sup>	Residual	$h^2$	$r$
1	1	56.1	68.5	1003	.050	.110
1	3	47.5	48.9	1028	.042	.086
1	5	47.4	76.4	968	.044	.114
3	1	47.3	58.8	986	.043	.097
3	3	40.3	45.7	821	.044	.095
3	5	52.9	36.6	901	.053	.090
5	1	39.3	48.4	748	.047	.105
5	3	31.4	43.1	658	.043	.102
5	5	36.0	58.0	829	.039	.102

<sup>A</sup>Permanent environment

High NGL% and low SDGL characterizes herds with more accurate reporting of insemination and calving dates. Previously, we found that heritability of GL was 14% when data from all herds was included. Imposing progressively higher quality standards on the data resulted in a progressively increasing heritability of GL. When the 10% of herds with lowest SDGL and highest NGL% were analyzed, heritability of GL increased to 21% (Zhang and Shook, 2001). Clearly, herd selection criteria based on GL did not improve heritability or repeatability of the economically important trait, DFB, in this study.

### CONCLUSIONS

Despite it's more nearly Normal frequency distribution, the log transformation of DFB did not result in increased heritability or repeatability. Thus it can be recommended to compute genetic evaluations for DFB as observed. Heritability or repeatability did not change with NGL% or SDGL. No advantage in accuracy of genetic evaluation would be gained by limiting genetic evaluations for DFB to herds with high NGL% or low SDGL.

### REFERENCES

- Hayes, J. F., Cue, R. I. and Monardes, H. G. (1992) *J. Dairy Sci.* **75** : 1701-1706.  
 Misztal, I. (2000) [ftp://num.ads.uga.edu/pub/blupf90/docs/remlf90.pdf](http://num.ads.uga.edu/pub/blupf90/docs/remlf90.pdf).  
 Pryce, J. E., Veerkamp, R. F., Thompson, R., Hill, W. G. and Simm, G. (1997) *Animal Sci.* **65** : 353-360.  
 Pryce, J. E., Coffey, M. P. and Simm, G. (2001) *J. Dairy Sci.* **84** : 1508-1515.  
 Thaller, G. (1998) Bulletin **18**, p. 55-61, International Bull Evaluation Service, Uppsala, Sweden.  
 Weigel, K. A. and Rekaya, R. (2000) *J. Dairy Sci.* **83** : 1072-1080.  
 Zhang, W. and Shook, G. E. (2000) *J. Dairy Sci.* **83** (Suppl. 1) : 53 [Abstract].  
 Zhang, W. and Shook, G. E. (2001) *J. Dairy Sci.* **84** (Suppl. 1) : 410 [Abstract].