

Merit of Outliers for Milk Yield as Indicators of Accuracy of Genetic Evaluations of Sires

T. R. MEINERT*¹ and H. D. NORMANT²

*National Dairy Herd Improvement Association, Inc.,
Columbus, OH 43231-4078

[†]Animal Improvement Programs Laboratory,
Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

To determine whether the accuracy of the genetic evaluations of sires can be assessed by the presence of extreme daughter records, we studied herd-years with records from first-crop daughters of 217 Holstein bulls that were not sampled by artificial insemination (AI) organizations but that later entered AI. The presence of outliers for standardized milk yield was determined within herd-year. Outliers were defined as records exceeding 1.5 interquartile ranges below the 25th percentile or above the 75th percentile. Herd-years were separated into two groups based on whether or not an outlier daughter record was present for an AI bull that had initially been sampled through non-AI. Herd-years without daughter outliers from those bulls were divided into herd-years with 1) no daughter outliers from any bull, 2) only negative daughter outliers from other bulls, 3) only positive daughter outliers from other bulls, or 4) negative and positive daughter outliers from other bulls. Herd-years with daughter outliers from AI bulls initially sampled through non-AI were divided into herd-years with 1) only negative daughter outliers, 2) only positive daughter outliers, 3) positive daughter outliers from those bulls and negative daughter outliers from other bulls, or 4) both negative and positive daughter outliers. The relationship between the frequency of outlier classes and a change in the Modified Contemporary Comparison genetic evaluations (the difference between the last available second-crop evaluation and the next to the last first-crop evaluation) was examined with logistic regression. For AI bulls that were initially sampled through non-AI and having evaluations that decreased ≥ 386 kg, 9% of herd-years had positive first-crop daughter outliers and negative daughter outliers from other bulls; 38% had no out-

liers. For bulls with evaluations that increased ≥ 194 kg, comparable percentages were 2 and 53%.

(**Key words:** genetic evaluation, outlier, herd profile, non-AI sampling)

Abbreviation key: IQR = interquartile range, MCC = Modified Contemporary Comparison.

INTRODUCTION

The existence of a comprehensive program in the US for dairy records management has been an important component contributing to genetic improvement. From the initiation of a cow testing association in Michigan in 1906 (10), the US dairy record-keeping program grew steadily, relied on considerable support from university and extension personnel, and became the large farmer-managed DHI organization. As of January 1, 1997, 48% of the US milking herd was enrolled in a DHI test plan (11).

Uniformity was a hallmark of the DHI program in the US for many years because of the existence of a national set of rules (10). Tested cows were usually milk recorded and sampled monthly. Today DHI is more market oriented: the number of tests is more variable, and dairy producers have more control over the frequency of milk weight collection and component sampling. The emphasis is on uniform operating procedures and documentation of testing rather than on rigid rule enforcement (4).

A program has been developed that will provide more information about characteristics of the herd so that industry users can better assess the accuracy of individual herd data. A uniform herd summary page (herd profile) was designed to present information graphically and was implemented (3) so that users of DHI records could draw conclusions about a herd quickly. The Council on Dairy Cattle Breeding mandated that dairy record processing centers develop the capability to provide herd profiles to end users with a fee per request by January 1997. Diagrams on the herd profile page clearly illustrate the testing situation.

Received November 21, 1997.

Accepted August 13, 1998.

¹Current address: Central Research Division, Pfizer, Inc., Eastern Point Road, Groton, CT 06340.

²Reprint requests.

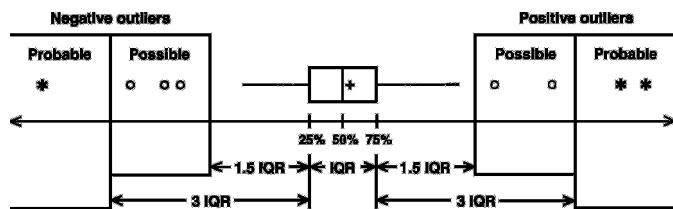


Figure 1. Example box plot (IQR = interquartile range).

A box plot is a graphical display of data that shows the location, spread, skewness, length of tails, and outlying data points (8). An example box plot is shown in Figure 1. The lower end (left side) of the center box depicts the 25th percentile, and the upper end (right side) depicts the 75th percentile. The interquartile range (IQR) is a measure of the range of data and is the difference between the 25th and 75th percentiles. The median (50th percentile) is shown as a central vertical line that splits the box into two parts, and the mean is marked with a plus sign (+). The dashed lines extending from the ends of the box are its "whiskers". The end of the left whisker marks the data point with the lowest value within 1.5 IQR below the 25th percentile. Similarly, the end of the right whisker designates the data point with the highest value within 1.5 IQR above the 75th percentile. Outliers are observations that fall outside the whiskers. Outliers are further classified as possible or probable. Possible outliers are between 1.5 and 3.0 IQR beyond the ends of the box (25th and 75th percentiles) and are data that may be unusual. Probable outliers are data that are more extreme and unusual than those designated as possible outliers; their values are >3 IQR beyond the ends of the box. The presence of probable outliers is extremely rare in a normal density function. Therefore, if outliers occur frequently, further scrutiny of the data may be needed to determine the cause.

One important component of the herd profile is a box plot for standardized milk yield that reveals the number and identification of individual cows with unusual deviations. In a herd environment that is free from preferential treatment, standardized milk yield is likely to have a distribution that resembles a normal density function because several of the components of records (e.g., interaction of herd and sire and effect of permanent environment) are random variables and, therefore, usually are not considered predictable. A large number of cows with extreme deviations for standardized yields within a herd-year likely indicate that an external source of variation is influencing the records without consideration of the

source. Nevertheless, an animal with an extreme deviation should not be assumed to have been subjected to preferential treatment because extreme deviations are expected occasionally from the normal density function as well as from most other distribution functions. However, if animals do receive extensive preferential treatment, their records commonly have extreme deviations and are classified as outliers; therefore, the occurrence of an unusually high number of outliers within any herd-year should raise concern about both the accuracy and the use of the records from that herd-year.

A second trait displayed as a box plot on the herd profile is cow PTA for milk yield. Because of the high financial worth of cows with outstanding genetic merit, a number of breeders assemble one or, sometimes, many cows with extremely high PTA, and the breeders usually have considerably more interest in those animals than in others in the herd. Frequently, the owners continue to give those animals special attention, especially when choosing service sires. The mating of cows with high PTA to bulls with high PTA perpetuates extremely high PTA for progeny. Because those progeny often have substantially higher PTA than PTA of other cows in the herd, cow PTA are not expected to have a normal distribution, and herd-years with outliers for cow PTA are not expected to be useful for assessing the accuracy of sire evaluations. However, standardized milk yield, which has a distribution that is approximately normal within a herd, might reveal differences in the accuracy of yield data between herds.

Meinert and Pearson (2) showed that bulls that were not AI-sampled but were brought into AI service had evaluations that declined more than those from AI-sampled bulls after their second-crop daughters began lactating. Meinert and Pearson (2) indicated that first-crop evaluations were accurate indicators of the genetic merit for AI-sampled bulls, but not for non-AI-sampled bulls, and the initial daughters of some non-AI-sampled bulls appear to have received preferential treatment. This favored status appears to have caused non-AI-sampled bulls to be overestimated for genetic merit during the period when decisions were being made by AI organizations on purchasing or leasing such bulls. Thus, the herd-years in which these daughters were lactating contained information of varying accuracy, and such a data set could possibly be useful for examining the effectiveness of outlier tests in identifying future changes in genetic evaluations. The purpose of this study was to determine the usefulness of outlier diagnostics in detection of herd-years with potential bias and to assess the

merit of outliers for standardized milk yield as indicators of accuracy of genetic evaluations of sires.

MATERIALS AND METHODS

All Modified Contemporary Comparison (MCC) evaluations (1) for non-AI-sampled Holstein bulls that later entered AI service and had a second-crop evaluation with a repeatability of at least 0.90 were extracted from the USDA evaluation history tape. All edits, determination of non-AI-sampling status, and designation of first-crop versus second-crop evaluations were determined by the criteria of Meinert and Pearson (2). Bulls were born from 1970 to 1981. Bulls were defined as non-AI-sampled if their initial evaluation occurred <2 yr after the birth of their first AI daughter. For each of 217 bulls, the last available second-crop evaluation (repeatability ≥ 0.90) minus the next to the last first-crop evaluation was calculated as a measure of change in evaluation and was used as the independent variable in the logistic regression analysis. The final MCC evaluation was in January 1989. The second to the last first-crop evaluation was used rather than the last to avoid including any second-crop daughters. The MCC evaluations were used because only a limited number of non-AI bulls entering AI based on animal model evaluations had sufficient time for a second-crop evaluation and also because the group of bulls that were selected using MCC had biased first-crop evaluations (2).

Standardized lactation milk yields (5) were obtained for the daughters included in the first-crop evaluation of each bull. First-crop daughters of a bull were all daughters that were born prior to 9 mo after the first evaluation of the bull. The lactation records of these daughters were used to identify all herd-years in which first-crop daughters of non-AI-sampled bulls produced milk. The standardized lactation milk yields for all cows in these herd-years were obtained from the USDA-DHIA lactation database.

Univariate statistics of standardized lactation milk yields from all cows in herd-years with first-crop daughter information were computed for each herd-year using the SAS PROC UNIVARIATE procedure (6). Those statistics were used to determine which herd-years contained possible and probable outliers. Yield records of first-crop daughters of non-AI-sampled bulls were matched with those herd-years that had outliers to determine whether or not the outlier for that herd-year was from such a daughter. To reduce the number of outlier classes that could be defined, possible and probable negative outliers

TABLE 1. Number of possible and probable outliers expected by herd size.

Herd size	Negative outliers expected		Positive outliers expected	
	Possible	Probable	Possible	Probable
	(no.)			
<142	0	0	0	0
287	1	0	1	0
573	2	0	2	0
860	3	0	3	0
1147	4	0	4	0
1433	5	0	5	0
2867	10	0	10	0
-854,000	2979	1	2979	1

were combined, and possible and probable positive outliers were combined.

Herd-years were grouped into eight outlier classes. These eight classes were of two types: type a, the first-crop daughter of a non-AI-sampled bull was not an outlier for standardized milk yield, or type b, the first-crop daughter of a non-AI-sampled bull was an outlier for standardized milk yield. There were four type a classes based on the presence and type of outliers for standardized milk yield in the herd-year: 1) no outliers, 2) only negative outliers, 3) only positive outliers, and 4) both negative and positive outliers. There were four type b classes based on the presence of and kind of outliers for standardized milk yield in the herd-year: 1) only negative outliers (daughter of non-AI-sampled bull is a negative outlier); 2) only positive outliers (daughter of non-AI-sampled bull is a positive outlier); 3) daughter of non-AI-sampled bull is a positive outlier, and daughters of other bulls are negative outliers; and 4) all other combinations of outliers (daughter of non-AI-sampled bull is a negative outlier, and daughters of other bulls are positive outliers or both negative and positive outliers, or daughter of non-AI-sampled bull is a positive outlier, and daughters of other bulls are both negative and positive outliers).

For each bull, the frequency distributions of herd-year outlier classes were calculated. Logistic regression, utilizing the SAS PROC CATMOD procedure (7), was used to examine the relationship between the frequency of outlier classes and the change in genetic evaluations (as shown in Model [1]). PROC CATMOD was run twice with different constraints to obtain estimates for all classes.

To examine further the effect of the frequency of outlier classes on changes in evaluations, bulls were stratified into seven classes based on the amount of change in evaluation for milk. Then, the frequency distributions of herd-year outlier classes were ex-

TABLE 2. Regression equations for predicting the probability of occurrence of outliers in herd-years that had first-crop daughters of non-AI-sampled bulls that entered AI from the amount of change in their genetic evaluations.

Outlier class	Intercept (%)	Change in bull evaluation	
		Regression coefficient (%/kg)	P
First-crop daughter not an outlier			
No outliers	45	0.0001770	0.0001
Only negative outliers	21	-0.0000700	0.0001
Only positive outliers	12	-0.0000500	0.0001
Negative and positive outliers	15	-0.0000500	0.0001
First-crop daughter an outlier			
Only negative outliers	3	-0.0000008	0.7710
Only positive outliers	2	-0.0000016	0.5346
Positive outliers with daughters of other bulls as negative outliers	2	0.0000063	0.0122
Other combinations of outliers	1	-0.0000056	0.0059

mined by evaluation change class. The model used for logistic regression was

$$g[\text{Pr}(Y_i = i | x)] = \alpha_i + \beta'x \quad [1]$$

where Y = herd-year of first-crop daughter of non-AI-sampled bull, x = change in genetic evaluation of the non-AI-sampled bull, $g[\text{Pr}(Y = i | x)]$ = joint probability that Y is in the outlier class i given x , α = intercept for outlier class, and β' = vector of slope parameters.

RESULTS

The box plot analysis and the designation of outliers can be helpful in situations in which extreme deviations occur more frequently than expected. The procedure uses the IQR, which is derived from observations near the mean and is therefore not affected extensively by the extreme observations as is the variance or standard deviation. Table 1 is presented to show the expected frequency of outliers. The expected number of outliers was calculated using the z -table, which produces probabilities for a normal distribution. A possible positive outlier and a possible negative outlier should appear once for every 287 lactations. For herds to obtain two or three positive outliers, they should have close to 573 and 860 lactations completed, respectively. Probable outliers should appear much less frequently. A positive outlier and a negative outlier are expected once every 854,000 lactations; therefore, if observations follow a normal density function, even one probable outlier in a herd should be extremely rare, and multiple probable outliers would be even less likely. However, if the expectation of occurrence of possible outliers is calculated and printed, as is being done in the herd profile, some modification (relaxation) of these probabilities seems appropriate because unknown environmental sources of variation do occur that are not accounted for in the adjustment for standardizing yield.

Table 2 shows the intercepts and regression coefficients for predicting the change in Predicted Difference milk between the last available second-crop evaluation and the next to the last first-crop evaluation using the MCC procedure (1). This analysis was done to determine whether outlier classes existed that

TABLE 3. Outlier frequency in herd-years that had first-crop daughters of non-AI-sampled bulls that entered AI according to the amount of change in genetic evaluations of the non-AI-sampled bulls.

Change in bull evaluation ¹ (kg)	Bulls (no.)	First-crop daughter not an outlier				First-crop daughter an outlier			
		No outliers	Negative outliers	Positive outliers	Both outliers ²	Negative outliers	Positive outliers	Positive outliers with daughters of other bulls as negative outliers	Other outlier combinations
<-579	9	41	10	14	17	<1	4	12	2
-579 to -386	9	35	19	15	18	2	1	6	3
-385 to -193	53	37	16	12	21	8	3	3	1
-192 to -1	79	34	18	14	23	5	2	3	1
0 to 193	48	42	16	12	16	4	5	3	2
194 to 386	18	52	16	9	10	5	5	2	1
>386	1	71	6	6	1	0	15	1	1
All bulls	217	38	17	12	20	5	3	3	1

¹Difference between the last available Modified Contemporary Comparison genetic evaluation and the next to the last first-crop evaluation.

²Both positive and negative outliers.

had significance in predicting whether bull evaluations increase or decrease. All intercepts were significant at $P < 0.0001$ (not presented in the table), and five of the eight regression coefficients were significant at $P < 0.01$. Although this analysis was based on Predicted Difference milk from the previous genetic evaluation procedure (1), a similar examination would probably be equally useful in the current procedure (12) because of the high correlations between evaluations from the two methods (9). These regression coefficients could be used to predict which bull will have future decreases in PTA through an examination of the herd-years in which daughters of each bull are lactating.

Because of the significance of the regression coefficients, the frequency of occurrence of different kinds of outliers was determined for the bulls according to seven classes representing differing amounts of change in their genetic evaluations for milk. These were the same classes that were defined by Meinert and Pearson (2). Table 3 shows the percentage of herd-years falling into each type of outlier class averaged across bulls. For bulls with evaluations that decreased ≥ 386 kg, 9% of herd-years had positive outliers for first-crop daughters and negative outliers for daughters from other bulls; 38% had no outliers. For bulls with evaluations that increased ≥ 194 kg, comparable percentages were 2 and 53%. The bull with the largest evaluation increase (>386 kg) had an extremely high percentage (71%) of herd-years when no outliers occurred among his own daughters or among daughters of herd-mate sires. Bulls with evaluations that decreased had a higher frequency of first-crop daughters in herd-years with outliers for standardized milk yield. Of particular interest was that 12% of the first-crop daughters of bulls with the largest decreases in evaluations (>579 kg) were in herd-years in which one or more of these daughters were positive outliers and the other bulls with daughters in these same herd-years had outliers, but these were exclusively negative. This result compares to only 3% across all change classes.

CONCLUSIONS

Information from the box plots of standardized milk yield can help explain the accuracy of bull evaluations by an examination of the herd-year data. Bulls with evaluation decreases had a higher frequency of first-crop daughters in herd-years with outliers for standardized milk yield. Also, bulls with evaluation decreases had a higher frequency of herd-years for which first-crop daughters were positive outliers and

daughters of other bulls were negative outliers for standardized milk yield. As a minimum, the number of outliers should be considered by AI organizations before non-AI-sampled bulls are purchased. It seems likely that this information could be helpful in assessing other bulls sampled as well. Research is needed to determine how best to utilize this information in the current procedure for genetic evaluation. Additional research is needed also to determine whether and how information from box plots is applicable to assessment of the accuracy of genetic evaluations for cows.

ACKNOWLEDGMENTS

Genetic evaluations and pedigree data were available through industry participants in the National Cooperative Dairy Herd Improvement Program. Special thanks are extended to R. E. Pearson and M. M. Schutz for constructive suggestions during preparation of this manuscript.

REFERENCES

- Dickinson, F. N., H. D. Norman, R. L. Powell, L. G. Waite, and B. T. McDaniel. 1976. Procedures used to calculate the USDA-DHIA Modified Contemporary Comparison. USDA Prod. Res. Rep. 165:18-34. ARS, USDA, Hyattsville, MD.
- Meinert, T. R., and R. E. Pearson. 1992. Stability of evaluations of bulls sampled by artificial insemination and other organizations. *J. Dairy Sci.* 75:564-571.
- National Dairy Herd Improvement Association. 1997. Herd profile and description. (<http://www.dhia.org/lib04.htm>. Accessed Mar. 4, 1998).
- National Dairy Herd Improvement Association. 1997. National Dairy Herd Improvement Program uniform operating procedures effective January 1, 1997. (<http://www.dhia.org/lib02.htm>. Accessed Mar. 4, 1998.)
- Norman, H. D., P. D. Miller, B. T. McDaniel, F. N. Dickinson, and C. R. Henderson. 1974. USDA-DHIA factors for standardizing 305-day lactation records for age and month of calving. USDA ARS-NE-40. ARS, USDA, Beltsville, MD.
- SAS® Procedures Guide, Version 6, Third Edition. 1990. SAS Inst., Inc., Cary, NC.
- SAS/STAT® Users Guide, Version 6, Fourth Edition. 1990. SAS Inst., Inc., Cary, NC.
- Tukey, J. W. 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
- VanRaden, P. M., C. A. Ernst, G. R. Wiggans, and H. D. Norman. 1989. Ability of animal model and Modified Contemporary Comparison evaluations of parents to predict future progeny performance. *J. Dairy Sci.* 72(Suppl. 1):71.(Abstr.)
- Voelker, D. E. 1985. History of dairy recordkeeping. *Natl. Coop. DHI Progr. Handbook, Fact Sheet A-2*. Ext. Serv., USDA, Washington, DC.
- Wiggans, G. R. 1997. NCDHIP participation as of January 1, 1997. *Natl. Coop. DHI Progr. Handbook, Fact Sheet K-1*. Ext. Serv., USDA, Washington, DC.
- Wiggans, G. R., and P. M. VanRaden. 1989. USDA-DHIA animal model genetic evaluations. *Natl. Coop. DHI Progr. Handbook, Fact Sheet H-2*. Ext. Serv., USDA, Washington, DC.