A Method for Verifying Genetic Evaluation Results

***Bert Klei*[1]*, Thomas Mark*[2]*, Freddy Fikse*[2]*, and Tom Lawlor*[1]**
[1]*Holstein USA, Brattleboro, United States*
[2]*Interbull Centre, Uppsala, Sweden*

Introduction

The accuracy and unbiasedness (quality) of international genetic evaluations are dependent on the quality of the data submitted. The Interbull Centre conducts a number of population based initial data checks on each submitted evaluations. These include, basic statistics, comparisons with previous evaluations, sire variances, and checking for trends in mendelian sampling (**MS**). These checks also look at sub-groups or individuals.

When inconsistencies are observed, evaluation centers are asked to explain the reasons for this. Depending on the explanation, cause, and time available, either the submitted national evaluation can be used, a corrected one can be submitted, or it can be decided to use data from the previous routine evaluation. Even though checks are in place, situations can occur in which inconsistent data does not get noticed. This can have a major detrimental impacts on international evaluation results.

MACE are a combination of parent averages, MS, within country variances and correlations. Ideally, data verification methods should cover each of these components. This paper will describe a method that can be used to increase the error detection abilities of the Interbull Centre as well as the ability to pinpoint problem animals. The method uses evaluations and reliabilities from consecutive evaluations to determine whether observed changes in individual evaluations are within expectations given the change in reliability.

Material and Methods

Distribution of the difference of two consecutive evaluations

Evaluations and their reliabilities from two consecutive evaluations are used to determine whether observed changes in an evaluation are reasonable given the change in reliability. In other words, if one has evaluations (u) for a bull (b) at time t and t+1 one whishes to determine the expected distribution of $\hat{u}_{b(t+1)} - \hat{u}_{bt}$. When the two evaluations are estimated using BLUP, the expectation of the difference is zero and the variance of the difference is:

$$\operatorname{var}\left(\hat{u}_{b(t+1)} - \hat{u}_{bt}\right) = \operatorname{var}\left(\hat{u}_{b(t+1)}\right) + \operatorname{var}\left(\hat{u}_{bt}\right) - 2\operatorname{cov}\left(\hat{u}_{b(t+1)}, \hat{u}_{bt}\right)$$

Van der Werf *et al.* (1994) argued that $\operatorname{cov}\left(\hat{u}_{b(t+1)}, \hat{u}_{bt}\right) = \operatorname{var}\left(\hat{u}_{bt}\right)$ and therefore, $\operatorname{var}\left(\hat{u}_{b(t+1)} - \hat{u}_{bt}\right) = \operatorname{var}\left(\hat{u}_{b(t+1)}\right) - \operatorname{var}\left(\hat{u}_{bt}\right)$.

By using the relationship $\operatorname{var}(\hat{u}) = \operatorname{var}(u) - PEV(\hat{u})$, and assuming that $\operatorname{var}(u)$ is constant over time, the above equation reduces to:

$$\operatorname{var}\left(\hat{u}_{b(t+1)} - \hat{u}_{bt}\right) = PEV\left(\hat{u}_{bt}\right) - PEV\left(\hat{u}_{b(t+1)}\right).$$

This says that if there is no new information, an evaluation should not change. This makes sense if one has an accurate value for prediction error variance (PEV). As will be explained later the values for PEV in this study are based on an approximation based on effective daughter contributions (EDC, Fikse and Banos, 2001). To account for the approximation, the following equality was used for the covariance term instead,

$$\operatorname{cov}\left(\hat{u}_{b(t+1)}, \hat{u}_{bt}\right) = r_{pev} \times \operatorname{var}\left(\hat{u}_{bt}\right),\ r_{pev} \leq 1$$

This models the fact that there can be additional new information other than that contained in $EDC_{b(t+1)} - EDC_{bt}$ that influences the reliability of the evaluation at t+1. In this study a value of .95 was use for r_{pev}.

Some countries impose a time edit on the data going into the evaluations, others –re-estimate parameters. This can cause $EDC_{b(t+1)} < EDC_{bt}$, and as a result $\operatorname{var}\left(\hat{u}_{b(t+1)} - \hat{u}_{bt}\right)$ becomes negative. To take this into account, the variance of the difference for each bull was calculated as,

$$\mathrm{var}\left(\hat{u}_{b(t+1)} - \hat{u}_{bt}\right) = abs\left(PEV\left(\hat{u}_{bt}\right) - PEV\left(\hat{u}_{b(t+1)}\right)\right) + \left(2 - 2r_{pev}\right)\left(1 - \min\left(rel_{bt}, rel_{b(t+1)}\right)\right)\mathrm{var}(u)$$

Calculating Prediction Error Variance

The evaluation submitted to the Interbull Centre contains no information on PEV. Therefore the reliability is used to calculate this value,

$$PEV\left(\hat{u}_{bi}\right) = \left(1 - rel_{bi}\right)\mathrm{var}(u)$$

The reliability submitted has only two significant digits. This is not adequate for the scope of this project. In addition, reliabilities are approximated in various manners in the different countries. As a result it was decided to calculate the reliability from the EDC submitted using,

$$rel_{bi} = EDC_{bi} / \left(EDC_{bi} + \alpha\right)$$

where $\alpha = \left(4 - h^2\right)/h^2$

and subsequently use this to calculate PEV for each of the evaluations.

Estimating $\mathrm{var}(u)$

Ideally, an estimate for the within country sire variance should be based on the deregressed evaluations and not the sire evaluations. Since, estimates based on this are not available until late in the MACE process, and one wants to verify the data as soon as possible, an approximation was used. First, an estimate for the within country sire variance was obtained by calculating the variance of the submitted evaluation by birth year. The variance in birth year *y* was subsequently adjusted for the average reliability in that birth year (Calo *et al*., 1973). A pooled estimate for the variance was then obtained by weighing each birth year estimate by the number of bulls in the birth year. Birth years with less than 10 bulls were eliminated from the calculation for the pooled estimate. Finally, the estimate of $\mathrm{var}(u)$ was obtained by averaging the estimates from the two evaluations.

Data

Evaluations submitted for two subsequent Interbull evaluations from four countries were used in this study. These will be referred to as OLD and NEW. The four countries were chosen based on population size to illustrate the methodology.

Results and Discussion

Table 1 shows some basic statistics for the two evaluations. From this table it can be seen that country B has a more than desired change in mean and standard deviation while the mean for country D also changed more than expected. One acceptable reason for the results of country D could be a base change.

Table 2 shows the results for country B and D by birth year. Additional information on regressions and correlation are also shown. From this table it can be seen that for country B something is going on in 1988. The regression of the OLD evaluations given the NEW ones is only .56, while the correlation between the two sets of evaluations is only .76. All regressions are expected to be 1.00. In both cases the correlations for 1996 is lower than that for 1992. This is likely due to the fact that a significant number of bulls born in 1996 are still adding daughters to get their first crop evaluations. In addition the entries over all birth years illustrate that population parameters can hide problems.

Table 1. Basic statistics for the two evaluations for protein.

		Mean		St. Dev.		Rel. Change
Country	Bulls	OLD	NEW	OLD	NEW	St. Dev.
A	~200	8.2	8.0	14.4	14.7	-.02
B	~4000	15.0	14.3	15.7	14.6	.07
C	~7500	-9.2	-9.2	17.0	16.9	.00
D	~17,000	8.5	8.9	17.1	16.7	.02

Table 2. Statistics by selected birth years for protein.

Country B		Mean		St. Dev.		Regressions		
Birth Year	Bulls	OLD	NEW	OLD	NEW	NEW|OLD	OLD|NEW	Correlation
1988	317	5.1	0.4	17.4	12.9	.56	1.02	.76
1992	318	20.0	20.0	16.2	15.9	.98	.99	.98
1996	394	31.5	32.5	13.8	14.0	.98	.91	.95
All		*15.0*	*14.3*	*15.7*	*14.6*	*.89*	*1.00*	*.94*
Country D								
Birth Year								
1988	1371	-4.8	-4.5	17.8	17.3	1.00	1.00	1.00
1992	1622	13.9	14.0	17.2	17.0	1.00	.99	1.00
1996	1285	32.6	31.9	16.1	16.3	1.00	.93	.97
All		*8.5*	*8.9*	*17.1*	*16.7*	*1.00*	*.98*	*.99*

The type of proof field as required by Interbull allows for another way of breaking down the bulls in groups. Correlations between subsequent evaluation results for these groups are in Table 3. Expected values for the correlations (last column) are 1. However when bulls change type of proof, last column, this usually indicates that a bull added a large number of daughters to his evaluation and some changes and re-rankings can be expected. The message from this Table is that something is going on with the imported bulls in country B.

Table 4 shows the distribution of the standardized change versus the expected change for each of the countries. From this Table is that the problem in country B is associated with imported bulls (type of proof 21). Country A and B have more than 10 times the expected fraction in the extreme categories. For Country A the reason might be the small number of bulls in their population.

An additional check could be on the regression of the difference in evaluation on birth year. A regression different from zero suggests that different genetic trends are observed in the two evaluations, and hence a change in evaluation. For the four countries in this study the trend was .02, .37, -.05, -.05, respectively. This suggests that there was a change in trend in country B that needs to be explained.

Table 3. Correlation of evaluations by type of proof in OLD and NEW for protein.

country	11-11*	12-12	21-21	11-12
A	.99	.99		
B	.99	.99	.71	.
C	1.00	1.00		.93
D	1.00	1.00	.99	.98

* 11 – first crop daughters
12 – first and second crop daughters
21 – imported bull

Table 4. Expected and observed standardized change in protein evaluations.

Country	-∞ to –3.000	-3.000 to -1.645	-1.645 to 1.645	1.645 to 3.000	3.000 to ∞
Expected	.0013	.0487	.9000	.0487	.0013
A	.0175	.0263	.8947	.0395	.0219
B	.0385	.0532	.8516	.0488	.0079
C	.0016	.0134	.9770	.0072	.0008
D	.0018	.0140	.9702	.0125	.0014

Table 5 shows the results for Country B divided out by birth year and type of proof. Even though the distribution of the observed change is off across the board, the most notable differences are those for 1988 and the imported bulls (type of proof 21). The Table indicates that bulls born in 1988 have seen a decrease in their breeding values, while those born in 1996 appear to have gone up. Similarly, the imported bulls have gone down as well. This matches the increase in genetic trend that was pointed out in the previous paragraph.

Finally, a sample of bulls that changed by more than 15 units (lb PTA, kg BV) is listed in Table 6. This Table illustrates the influence of reliability change but also the influence of differences in variance estimates for each country on the test statistic (Criteria). Country A is not represented in this Table since no bulls in this country changed by more than 15 units. Bull 1 and 2 in country B see their evaluation change by a similar amount. Even though the second bull has a slightly less absolute change in evaluation than the first one, the fact that bull 1 added more information allowed it to change more. A similar situation occurs for both bulls in C. In this case bull I has a larger change in his reliability, and thus PEV, than bull II and is therefore allowed to change more. This is reflected in Criteria. Bull B in country D is a bull that might need some closer scrutiny. The value of the change Criteria is less than –3. In this case the explanation is that the bull almost tripled his number of EDC. The bull that really stands out in this group is bull 3 in Country B. Even though he added 53 EDC, this was only a tiny fraction when compared to the total number. As a result his evaluation was not expected to change, but he lost 29 units of protein. It turns out that he was a member of the group of bulls that was identified in Table 5 to likely have a problem.

Table 5. Expected and observed standardized change in protein evaluations by birth year and type of proof for Country B.

Country	-∞ to –3.000	-3.000 to -1.645	-1.645 to 1.645	1.645 to 3.000	3.000 to ∞
Expected	.0013	.0487	.9000	.0487	.0013
Birth Year					
1988	.1672	.0410	.7792	.0095	.0032
1992	.0094	.0629	.8396	.0818	.0063
1996	.0000	.0508	.8274	.1041	.0178
Type of Proof					
11	.0021	.0524	.8945	.0458	.0051
12	.0108	.0430	.9032	.0430	.0000
21	.3620	.0625	.4661	.0755	.0339

Table 6. Examples of changes observed in bull evaluation and the associated test criteria.

Country	Bull	Eval.	EDC	Calc. Rel.	Eval.	Criteria
B	1	OLD	12	.493	5	
		NEW	43	.778	24	2.167
	2	OLD	12	.493	33	
		NEW	21	.630	51	2.743
	3	OLD	1582	.992	29	
		NEW	1631	.992	0	-60.383
C	I	OLD	12	.493	21	
		NEW	44	.781	4	-1.718
	II	OLD	30	.709	14	
		NEW	44	.781	-2	-2.878
D	A	OLD	15	.549	35	
		NEW	46	.789	13	-2.441
	B	OLD	64	.838	40	
		NEW	150	.924	19	-3.899

A remaining question is when do results of these tests warrant closer scrutiny. It is obvious from this example that country B should take a closer look at its evaluation. In many cases discrepancies might not be as obvious.

Correlations in Table 2 less than .97, except for recent birth years, should require additional explanation. Similarly in Table 3, correlations for bulls staying in their type of proof sub-class should be around .97. For bulls changing sub-classes a correlation of .90 is reasonable. When looking at standardized change it is not expected that more than .2% of bulls are in each of the extreme categories. Some allowances should be given for small populations. Finally, it is also recommended that all bulls with an absolute value for the test criteria greater than 4 be printed out. This can then be used to further determine whether any systematic problems in the data occur. It can also be used to inform evaluation centers which bulls have shown large changes in their evaluations.

In the Introduction it was mentioned that ideally within country variances, correlations, PA and MS should be checked. This paper describes a more in depth verification process of within country variances and consistency of evaluations. It does not address the question of whether an evaluation in a country is correct. This could be addressed by examining the distribution of MS. Ideally, this would again be done before MACE calculations are started. However, data submitted to Interbull do not have all necessary information to effectively determine MS. Instead of using the submitted data one can use MACE results for this verification step. This method of data verification is currently being developed. This leaves open the question of validation of correlations.

Conclusion

The method and process described in this paper shows that differences in genetic evaluations combined with basic statistics can be used to identify problem evaluations. Instead of only identifying problems on the population level it will help find groups and individual animals whose evaluations justify closer scrutiny. The method can easily be implemented in each individual country as an additional check on genetic evaluation results before data is submitted to the Interbull Centre for routine evaluations.

Acknowledgements

The authors would like to thank all countries that have graciously allowed for the use of their data in this study.

Literature

Calo, L.L., McDowell, R.E., VanVleck, L.D. & Miller, P.D. 1973. Genetic aspects of beef production among Holstein-Friesians selected for milk production. *J. Anim. Sci. 37*, 676-682.

Fikse, W.F. & Banos, G. 2001. Weighting factors of sire daughter information in international genetic evaluations. *J. Dairy Sci. 84*, 1759-1767.

Van der Werf, J.H.J., Meuwissen, T.H.E. & de Jong, G. 1994. Effects of correction for heterogeneity of variance on bias and accuracy of breeding value estimation for Dutch dairy cattle. *J. Dairy Sci. 77*, 3174-3184.