THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Effect of time period of data used in international dairy sire evaluations

OPEN ACCESS

GENETICS AND BREEDING

Effect of Time Period of Data Used in International Dairy Sire Evaluations

K. A. WEIGEL* and G. BANOS†
*Department of Dairy Science, University of Wisconsin,
1675 Observatory Drive, Madison 53706
†INTERBULL Centre, Department of Animal Breeding and Genetics,
Swedish University of Agricultural Sciences, S-750 07 Uppsala, Sweden

ABSTRACT

Simulation was used to examine the changes that may occur in international dairy sire evaluations when differing amounts of historical performance data are used. When the base genetic variances within each population were equal, use of all historical data in international sire evaluations gave unbiased predictions for breeding values. When some historical data from one population were discarded, the estimated genetic standard deviation for this population was reduced because of the effects of genetic selection, and slope coefficients of conversion equations were biased slightly in favor of this population. When base genetic variances differed among populations, the use of all historical performance data in international sire evaluations resulted in substantial upward bias in evaluations of elite sires from the importing population. This result occurred because the estimated genetic standard deviation in the importing population was reduced by inclusion of performance data from many older bulls from the breed or strain that was being replaced. When performance data from bulls that were born prior to the beginning of importation were discarded, estimated parameters for genetic standard deviation were similar for both populations, and estimated breeding values of elite bulls were very close to the true values. This research suggests that accuracy of international dairy sire evaluations may be improved by discarding historical performance data of bulls from breeds or strains that have been replaced by imported stock.
(**Key words**: international comparisons, dairy sires, importation)

Abbreviation key: **MACE** = multiple-trait across country evaluation, **REL** = reliability, **SD** = standard deviation.

INTRODUCTION

International genetic evaluations of elite dairy sires calculated using the multiple-trait across country evaluation (**MACE**) procedure (1, 4) are sensitive to changes in parameters for genetic standard deviation (**SD**) within a country (5). The underestimation of genetic SD for a particular country can lead to the overevaluation of elite bulls from this country, and overestimation of genetic SD for another country can lead to underevaluation of elite bulls from that country. Estimated genetic SD within a country can depend on the time period of the data used for estimation because genetic SD parameters are affected by genetic selection, importation, the presence or absence of adjustments for heterogeneous variance, and other factors.

Selection generally reduces genetic variance (2); however, the variance for milk production traits tends to increase, because of scaling, as the mean increases (7). In practice, some countries have much more historical data available to account for the effects of genetic selection. For example, only 6% of Holstein bulls evaluated in Denmark were born before 1980, but 54% of Holstein bulls evaluated in Great Britain were born before 1980. The availability of such historical data often depends on factors other than genetics, such as the date by which the national dairy database was established.

Importation can substantially affect estimated genetic SD within a country, particularly if genes are imported from a breed or strain that has higher mean and variance. A. Koerhuis (1996, personal communication) found that the estimated genetic SD in The Netherlands was 6% larger for bulls of ≥62.5% North American Holstein inheritance than for a combined file including Holstein and Dutch Friesian bulls; this difference was 11% in the absence of an adjustment for heterogeneous variance. Although adjustment procedures for heterogeneous variance within national genetic evaluations can partially account for changes in genetic SD across time, differences exist across countries in the existence and methodology of procedures for standardization of within-herd variances (3).

Substantial differences in the beginning date and rate of importation exist across countries. For example, Schaeffer et al. (5) noted that 41% of French Holstein bulls born between 1970 and 1978 were of North American origin compared with 89% of bulls born after 1978. However, 84% of German Holstein bulls born between 1970 and 1978 were of North American origin compared with 97% of the bulls born after 1978 (5). Schaeffer et al. (5) noted that differences among European countries in the rate of adoption of North American genetics may necessitate limiting MACE evaluations to bulls born in the last 15 yr.

In practice, limiting milk production data used in international sire evaluations to bulls born in 1980 or later results in changes of up to 15% in estimated genetic SD within a country, which can significantly affect the international rankings of elite bulls. The objective of this study was to examine, using simulation, the changes in international EBV and rankings that occur when data from different time periods are used in national and international dairy sire evaluations.

MATERIALS AND METHODS

Simulation of Data

Breeding values for sires and cows in two hypothetical populations were simulated as bivariate normal; the genetic correlation between populations was 0.93. Phenotypic data were generated according to a heritability value of 0.30. In the control situation, the genetic mean and SD were assumed to be 100 and 10.0, respectively, in both populations. In the upgrading situation, the genetic mean and SD parameters were 100 and 10.0, respectively, in population A and 125 and 12.5, respectively, in population B. Therefore, population A could be considered as the importing population.

Eight generations of selection were carried out with 36,000 cows and 300 progeny test bulls in each population per generation. In each generation, 10 sires and 150 dams were chosen as parents of bulls for the next generation. In addition, 30 bulls were chosen based on progeny test results for use as sires of cows in the next generation. Each test bull had 80 progeny in the current generation, which is currently the goal of most US AI companies, and bulls that were selected as sires of cows had 400 additional progeny in the following generation.

Selection was carried out within populations in generations 1 through 3, and genetic exchange between populations was allowed in generations 4 through 8. A maximum of 50% of sires of sons, dams of sons, and sires of cows were selected from the other population in each of generations 4 through 8. Genetic and residual SD for an individual animal were calculated as a function of the base parameters for each population and the proportion of genes that had been inherited from each base population.

Genetic selection was based on true breeding value plus or minus a random error term. The magnitude of this random error term varied, depending on the desired reliability (**REL**) of the bull and cow selection in each generation. Accuracy of local bull selection (i.e., selection among bulls evaluated within the population) corresponded to a REL of 0.60 in generation 1, 0.625 in generation 2, 0.650 in generation 3, . . ., 0.80 in generation 8, which corresponds to a situation of improved accuracy of data collection or genetic evaluation methodologies over time. The accuracy of foreign bull selection (i.e., selection among bulls evaluated only in the other population) was $(0.93)^2$ times the accuracy of local bull selection, and the accuracy of cow selection was 0.66 times the accuracy of bull selection, so accuracy of both bull and cow selection depended on generation and population of origin. Ten replicates of each scenario were carried out.

EBV Within a National Population

Genetic evaluations within each population were calculated using a single-trait animal model that defined genetic groups for missing parents according to generation and population of origin. True heritability parameters were used in animal model calculations. Two scenarios were considered. In the first case, all available performance data (generations 1 to 8) were used to calculate national EBV. In the second case, performance data from cows born in generation 3 to the present were used in national EBV in population A, and all performance data (generations 1 to 8) were used in national EBV in population B. All historical data for pedigree of sires were retained in both populations. This situation corresponds to a situation in which an importing country may have implemented its national data recording or genetic evaluation system later than an exporting country.

International EBV

International sire evaluations were calculated using the MACE procedure of Schaeffer (4). Genetic SD were estimated for each country using the approximate REML procedure of Sigurdsson and Banos (6),

TABLE 1. Estimated genetic standard deviation[1] of populations A and B, which were used in the multiple-trait across country evaluations (MACE).

Data	Control		Upgrading	
	A	B	A	B
All data used in national evaluations				
All data used in MACE	10.18	10.11	11.98	12.72
Partial data used in MACE	9.61	9.60	12.28	12.11
Partial data used in national evaluations				
All data used in MACE	9.86	. . .	12.14	. . .
Partial data used in MACE	9.61	. . .	12.28	. . .

[1]Mean of 10 replicates.

and true heritability parameters were used in genetic SD estimation. These estimated genetic SD were used with the true genetic correlation in the MACE analysis. Genetic groups for missing parents were defined according to generation and population of origin. Second-country data of bulls that were first evaluated in the other population were included in estimation of genetic SD and MACE EBV. Two situations were considered regarding the selection of data for international evaluations. In the first case, all available national evaluation data were used in the MACE analysis, which included either data from generations 1 to 8 in both populations or data from generations 3 to 8 in population A and generations 1 to 8 in population B, depending on which data were used in the national evaluations for each population. In the second case, national evaluation data of both populations were edited to include only those bulls that were born in generation 4 or later; however, all pedigree data from the older bulls were retained. In this case, the MACE analysis included data from the same time period for each country, regardless of the amount of historical data that had been included in each national evaluation.

RESULTS AND DISCUSSION

Control Situation

All phenotypic data used in national EBV. When base genetic mean and SD were the same in both populations and when all data were used in national evaluations, estimated genetic SD for populations A and B were similar, as shown in Table 1, regardless of which national evaluation data were used in the MACE analysis. Estimated genetic SD decreased for both populations when the MACE analysis was limited to bulls that were born in generation 4 or later because the effects of selection in early generations were ignored. However, no significant differences occurred in the bias of estimation (Table 2), international sire rankings (Table 3), or conversion equations (Table 4), although the absolute error of MACE prediction for elite bulls was smaller when the MACE analysis included only those bulls born in generation 4 or later (Table 5).

Phenotypic data from generation 3 or later used in national EBV in population A. When national evaluations in population A used only data

TABLE 2. Mean bias[1,2] in EBV from the multiple-trait across country evaluations (MACE).

Data	Control		Upgrading	
	All bulls	Top 10	All bulls	Top 10
All data used in national evaluations				
All data used in MACE	–0.1	+0.5	+2.6**	+3.6**
Partial data used in MACE	–0.4	+0.1	+0.9**	+0.1
Partial data used in national evaluations				
All data used in MACE	–0.3	+0.6	+1.5**	+2.1**
Partial data used in MACE	–0.4	+0.1	+0.8**	+0.1

[1]Bias for all generation 7 and 8 bulls and for the top 10 bulls was calculated as [(error for A bulls in B) – (error for B bulls in A)]/2.

[2]Mean of 10 replicates.

**Significantly different from 0 ($P < 0.01$).

TABLE 3. Number of bulls[1] from population A ranking among the top 100 bulls internationally on the scale of each population.

Data	Control		Upgrading	
	A	B	A	B
True representation	59.3	41.8	56.7	41.6
All data used in national evaluations				
All data used in MACE[2]	62.4	39.6	70.3**	51.4**
Partial data used in MACE	59.0	41.1	58.9	43.2*
Partial data used in national evaluations				
All data used in MACE	62.7	40.0	66.3**	46.2**
Partial data used in MACE	59.0	41.1	58.9	43.2*

[1]Mean of 10 replicates.

[2]Multiple-trait across country evaluations.

*Significantly different from true representation ($P < 0.05$).

**Significantly different from true representation ($P < 0.01$).

from cows born in generation 3 or later, the estimated genetic SD for population A was slightly smaller than that for population B (Table 1). This difference occurred because the effects of genetic selection in the first three generations were taken into account in population B but were ignored in population A (for which phenotypic data were discarded). Proportions of bulls from each population ranking among the top 100 were not significantly different from true proportions (Table 3), and estimated bias in MACE predictions was not significantly different from zero (Table 2). However, slope coefficients of conversion equations were slightly inflated for conversions from population A to B and reduced for conversions from population B to A (Table 4); intercept coefficients also changed slightly to compensate for the differences in slope coefficients. The absolute error of MACE prediction for elite bulls was once again smaller when the MACE analysis included only those bulls born in generation 4 or later, as shown in Table 5.

Upgrading Situation

All phenotypic data used in national EBV. When base genetic mean and SD were different in each population, inclusion of all national evaluation data in MACE evaluations gave a substantially smaller estimated genetic SD for population A than for population B (Table 1). This difference occurred because many older bulls from base population A were included in the estimation of the genetic SD for population A, and these bulls had lower mean values and less variance than the imported bulls. These population A bulls from generations 0 to 3 had very few genetic relationships with current population A bulls, which were mostly descendants of imported animals. When the MACE analysis was limited to bulls that were born in generation 4 or later, SD parameters were very similar for both populations.

When all national evaluation data were used in the MACE analysis, MACE predictions for average and elite bulls were significantly biased in favor of the

TABLE 4. Conversion formulas[1] calculated from regression of solutions of multiple-trait across country evaluations (MACE) in the importing country on MACE solutions in the exporting country for bulls born in generations 7 and 8.

Data	Control		Upgrading	
	A to B	B to A	A to B	B to A
All data used in national evaluations				
All data used in MACE	$13.9 + 0.921X$[2]	$14.4 + 0.920X$	$8.5 + 0.986X$	$28.3 + 0.848X$
Partial data used in MACE	$12.7 + 0.928X$	$14.7 + 0.923X$	$19.9 + 0.933X$	$13.2 + 0.922X$
Partial data used in national evaluations				
All data used in MACE	$9.4 + 0.943X$	$18.9 + 0.899X$	$9.6 + 0.976X$	$26.2 + 0.862X$
Partial data used in MACE	$12.7 + 0.928X$	$14.7 + 0.922X$	$19.9 + 0.933X$	$13.2 + 0.922X$

[1]Mean of 10 replicates.

[2]X = Genetic evaluation in the exporting country.

TABLE 5. Mean absolute error[1,2] in EBV from the multiple-trait across country evaluations (MACE).

	Control		Upgrading	
Data	All bulls	Top 10	All bulls	Top 10
All data used in national evaluations				
All data used in MACE	3.9	5.3	5.2	6.8
Partial data used in MACE	3.8	4.4**	4.6*	4.9**
Partial data used in national evaluations				
All data used in MACE	3.9	5.2	4.8	6.2
Partial data used in MACE	3.8	4.4**	4.6	4.9**

[1]Absolute error for all generations 7 and 8 bulls and for the top 10 bulls was calculated as [(error for A bulls in B) + (error for B bulls in A)]/2.

[2]Mean of 10 replicates.

*Significantly less than error with all data used in MACE ($P < 0.05$).

**Significantly less than error with all data used in MACE ($P < 0.01$).

importing population (Table 2). In addition, MACE rankings on the scale of population A had an extra 13.6 bulls from population A among the top 100 bulls compared with true values (Table 3). The MACE rankings on the scale of population B had an extra 9.8 bulls from population A among the top 100 bulls compared with true values. Finally, the slope coefficients of the conversion equations to and from population A were biased in favor of population A bulls (Table 4). Therefore, the genetic SD for population A, which was estimated from a mixture of local and imported bulls, was apparently not appropriate for standardizing national EBV of the current population A bulls, which contained nearly 100% foreign (population B) genes.

When the MACE analysis was limited to bulls born in generation 4 or later, mean bias (Table 2) and mean absolute error of MACE prediction (Table 5) were substantially reduced, particularly for elite bulls. The number of bulls from each population ranking among the top 100 on the scale of population A were not significantly different from true values (Table 3); the bias in rankings was also greatly reduced on the scale of population B, although the small deviation from true values was still marginally significant. Conversion equations between the two populations had slope and intercept coefficients that were similar in magnitude (Table 4).

Phenotypic data from generation 3 or later used in national EBV in population A. When national evaluations from population A were restricted to included data from cows born in generation 3 to the present, the estimated genetic SD for population A were slightly larger than those obtained when all data were used in national evaluations (Table 1), but these values were still much smaller than the

estimates that were obtained using all data for population B.

When all national evaluation data were used in the MACE analysis, MACE predictions for average and elite bulls were again significantly biased in favor of the importing population (Table 2), but differences were slightly smaller than those that occurred when national evaluations within population A included all historical data. The MACE rankings on the scale of population A has an extra 9.6 bulls from population A among the top 100 bulls, compared with true values, and MACE rankings on the scale of population B had an extra 4.6 bulls from population A among the top 100 bulls, compared with true values (Table 3). The slope coefficients of the conversion equation to and from population A were again biased in favor of population A bulls (Table 5).

When the MACE analysis was limited to bulls born in generation 4 or later, mean bias and mean absolute error of MACE prediction were again reduced (Tables 2 and 5), and the number of bulls from each population ranking among the top 100 on the scale of population A were much closer to true values (Table 3). Conversion equations between the two populations had slope and intercept coefficients of similar magnitude (Table 4).

CONCLUSIONS

The time period of data used for national and international evaluations of dairy sires can have a significant impact on international sire rankings and conversions, particularly when some countries replace local cattle with imported stock of higher genetic merit. The effects of time period in international sire evaluations are primarily a reflection of changes in genetic SD within participating countries over time;

accurate estimates of genetic SD for each country are necessary to obtain correct international rankings of current elite sires. Genetic SD within a population can change over time because of genetic selection, importation, the presence or absence of heterogenous variance adjustments in national genetic evaluations, and other factors. Among countries that currently participate in international dairy sire evaluations, many differences exist in the amount of historical data that are stored and used in national genetic evaluations (and, hence, in the ability to account for selection effects on the variance). Differences also exist in the beginning date of importation, in the rate at which local stock are replaced with genes of another breed or strain, and in the structure and implementation of the national model for genetic evaluation.

This simulation study supports the hypothesis that accuracy and fairness of international dairy sire evaluations will be improved if these evaluations are based on recent national evaluation data (e.g., from bulls born within the last 15 yr). Dairy cattle populations are constantly changing across time, and the current cattle populations in some countries may bear little resemblance or relationship to those that were present 25 yr earlier. Use of more recent data should provide parameter estimates that are more appropriate for application to the current generation of elite bulls from each country. If a minimum year of birth (e.g., 1980) is implemented for international dairy sire evaluations, it will probably be necessary to update this criteria periodically because some countries are currently still in the process of upgrading with Holstein cattle. Furthermore, selection intensity is likely to differ within countries, and a constant base year for all countries may be an inappropriate assumption. Future technical developments could allow appropriate use of more historical data. For example, multiple-trait estimation of genetic variances within a country (when performance in each country is considered as a different trait) would allow selection effects on the genetic variance to be taken into account in both importing and exporting counties. In addition, adoption of more uniform procedures for heterogenous variance adjustment across counties or editing of data based on breed composition might allow historical performance data to be used more appropriately.

REFERENCES

1 Banos, G., and A. Sigurdsson. 1996. Application of contemporary methods for the use of international data in national genetic evaluations. J. Dairy Sci. 79:1117.

2 Bulmer, M. G. 1971. The effect of selection on genetic variability. Am. Nat. 105:201.

3 International Bull Evaluation Service. 1992. Sire Evaluation Procedures for Dairy Production Traits Practised in Various Countries. Bull. No. 5. INTERBULL, Uppsala, Sweden.

4 Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. J. Dairy Sci. 77:2671.

5 Schaeffer, L. R., R. Reents, and J. Jamrozik. 1996. Factors influencing international comparisons of dairy sires. J. Dairy Sci. 79:1108.

6 Sigurdsson, A., and G. Banos. 1995. Estimation of genetic correlations between countries. Page 1 in Proc. INTERBULL Annu. Mtg., Prague, Czech Rep. Bull. No. 11. INTERBULL, Uppsala, Sweden.

7 Weigel, K. A., D. Gianola, B. S. Yandell, and J. F. Keown. 1993. Identification of factors causing heterogeneous within-herd variance components using a structural model for variances. J. Dairy Sci. 76:1466.