Multiple-Country Comparison of Dairy Sires

L. R. SCHAEFFER
Department of Animal
and Poultry Science
University of Guelph
Guelph, ON, Canada N1G 2W1

ABSTRACT

A multiple-trait sire model was described in which yields of daughters in different countries were considered to be different traits. Such a model required estimates of genetic correlations among sire genetic effects in different countries. Observations were average daughter yield deviations, which were yields adjusted for all fixed effects within a country and for mate and animal permanent environmental effects. The methodology was described through a small example. Methods for estimating genetic covariances between countries as well as the advantages and disadvantages of the multiple-country approach to international comparisons were discussed. The proposed method appears to be better theoretically than methods in which the genetic correlations among countries are assumed to be 1. The relationships of estimated transmitting abilities of sires from the multiple-country analysis with international conversion formulas were discussed.

(Key words: international comparisons, multiple traits, sire models)

Abbreviation key: BCA = breed class average, DYD = mean daughter yield deviations, MGD = maternal granddam, MGS = maternal grandsire.

INTRODUCTION

The international comparison of dairy sires is an important issue to organizations that import and export semen, embryos, and cattle. INTERBULL (Uppsala, Sweden) was founded to standardize methodology used for making comparisons among countries. The simplest

comparison uses the regression of sire proofs from the importing country on those from the exporting country. INTERBULL has recommended minimum reliabilities for bull proofs to be included in the development of the regression equations and in the actual regression methodology (6). For the most part, countries appear to be complying with these recommendations.

In 1985, Schaeffer (10) proposed a linear model method to combine sire proofs from different countries into a single international proof. Rozzi et al. (9), Banos et al. (2), and Banos (1) used this procedure, with modifications, to compare populations of dairy cattle. The method assumes that there are no interactions of genotype and environment and that the genetic correlation between daughter yields in all pairs of countries is unity. The genetic evaluation methods of each country also are assumed to remove as much bias from nonrandom mating and preferential treatment as is technically possible. Another assumption is that heritability is the same for each country. Schaeffer (10) used bull proofs as observations in the model; Rozzi et al. (9) used deregressed bull proofs. Average daughter yield deviations (DYD) are now being used for obtaining conversion formulas (7). The DYD of a bull is the weighted average of mean of daughter yields deviated from solutions for the effects of herdyear-season, age and month of calving, interactions of herd and sire, genetic merit of mates, and cow permanent environment that is specific to each daughter (14). The DYD are analogous to deregressed bull proofs.

The purpose of this paper is to describe a multiple-trait sire model that treats the DYD of bulls in different countries as different traits. The calculations are described using a small example involving only two countries.

THE MODEL

The model that describes a bull's mean DYD from country i is

Received September 8, 1993. Accepted May 3, 1994.

$$y_i = \mu_i 1 + Z_i Q g_i + Z_i s_i + e_i$$

where

y_i = the vector of DYD from country i for one particular trait such as protein yield.

 μ_i = a scalar for country i, which reflects the definition of the genetic base for that country,

g_i = a vector of genetic group effects of phantom parents,

s_i = a vector of random sire transmitting abilities for country i,

 e_i = a vector of effects of random mean residuals,

 Z_i = the matrix that relates DYD to sires, and

Q = the matrix that relates sires to phantom parent groups.

For t countries, the variance-covariance matrices of the random vectors are

and

where s_{ij} is the sire (co)variance between countries i and j, A is the additive genetic relationship matrix for all bulls based on sire and maternal grandsire (MGS) relationships, σ_{ei}^2 is the residual variance for country i, and D_i is a diagonal matrix with elements equal to 1 over the number of daughters in a bull's DYD. The number of daughters should be directly related to the accuracy of a bull's proof, or the number

of daughters could be approximated from the accuracy figure, or the sum of the weights on DYD within a bull could be used. Covariances among DYD within a country are assumed to be zero, although in reality this assumption is not true. Ignoring the covariances should not bias results from the multiple-trait model but would result in predictions of sire effects having greater prediction error variances than if the covariances were taken into account. To incorporate the covariances would require the inverses of the coefficient matrices of Henderson's mixed model equations (6) from the animal models employed in each country. Thus, ignoring the covariances is a practical alternative.

This formulation of the model allows for different residual and sire variances for each country and, therefore, allows for different heritabilities for each country. The sire covariances also allow for genetic correlations between countries that are less than unity, which could reflect interactions of genotype and environment. Genetic correlations would also be <1 if the animal models for different countries differ in their ability to account for all nongenetic effects.

The multiple-trait model also allows the trait to be measured differently for each country. For example, yields in Canada are measured with breed class averages (BCA); yields are commonly expressed in kilograms in Europe and in pounds in the US. With the model of Schaeffer (10), the traits, which could be expressed as breeding values or transmitting abilities, need to be standardized within each country before being analyzed together, but with the multiple-trait model this standardization is not necessary.

The multiple-trait model and mixed model methodology for predicting genetic values of animals with data not in the model was first described by Henderson (4, 5), and this paper is an application of that methodology.

Genetic Groups

Banos et al. (2) defined genetic groups of sires on the basis of country of origin and year of birth so that the origins of blood lines for each bull must be determined prior to analysis. The origin was defined as one-half the origin of the sire plus one-quarter the origins of the

dam and MGS (1), which in essence followed the idea of Westell et al. (15) of phantom parent grouping. Thus, only unknown parents need to be assigned to genetic groups. For example, for a bull that is first registered in Canada with unknown parents, the unknown sire of that bull would be assigned to a sire phantom group corresponding to Canadian origin and year of birth of that bull. The unknown MGS would also be assigned to a MGS phantom group corresponding to Canadian origin and year of birth of that bull. Thus, unknown parents would be assigned to only one country of origin. Subsequent sons and grandsons, however, may have a mixture of blood lines from two or more countries. Phantom parent groups simplify the determination of country of origin. The groupings used in the multipletrait international model would be different from groupings used within each country for evaluations.

Connectedness

Each vector s_i has length equal to the total number of bulls in all t countries that have DYD plus ancestor bulls that may not have a DYD in any country. The relationship matrix, A, is assumed to provide ties among bulls from different countries and should be checked carefully prior to data analysis. Much work is needed to quantify the degree of connectedness among countries and the effects of low degrees of connectedness on international comparisons among multiple countries.

Preferential Treatment

Currently the computation of regression conversion equations for two countries is based on bulls that have daughters in both countries. The daughters in the importing country may be preferentially treated for several reasons, and thus their use for obtaining conversion formulas has been questioned (1). One recommendation for the linear model approach of Schaeffer (10) has been to not use proofs of imported bulls in the analysis (1). This recommendation appears to be appropriate and should also apply to the multiple-trait model presented here. Thus, only DYD of bulls from the country in which they were first progeny tested should be included. In the future, the distinction of where a bull is first progeny tested may become less clear. For example, some Canadian bulls now being progeny tested simultaneously in Canada and Europe will receive their first proof at approximately the same time in each country. In this situation, both proofs might be suitable for the multiple-trait model. Some Canadian bulls are also being progeny tested in Europe, but not in Canada. This problem should be alleviated by the grouping strategy for phantom parents.

Unique Identification

A practical problem for international comparisons is that a bull may have various registration numbers in different countries, and each number may be used frequently. An international format of 15 characters was adopted during 1992; the first 3 characters are a country code, and the last 12 are the registration number for the country of birth (INTERBULL Meeting, Neustift, Austria). This system has been adopted by a few countries for exchange of genetic information, but problems with pedigrees still exist because the sires and MGS have not been addressed as carefully as the bulls themselves. Thus, some bulls likely may receive more than one evaluation from an international multiple-trait model; that is, first, from having daughters in one country under one identification, and, second, from having sons or grandsons in another country under a second identification. INTERBULL is constructing crossreference tables among countries to tackle this problem.

EXAMPLE

Consider protein yield data on bulls from two countries, say AA and BB, with proofs expressed in different units of measurement, say kilograms and BCA, as shown in Table 1. In this example, one bull was progeny tested simultaneously in both countries, and other bulls had DYD in only one country. To analyze the data with the multiple-trait model, estimates of residual and sire variances and the genetic correlation between the two countries need to be available. Assume for this example that sire variances are 100 kg² and 5 BCA², that the sire covariance is 20-kg BCA, and that the residual variances are 1000 kg² and 80 BCA² for countries AA and BB, respectively.

2674 SCHAEFFER

TABLE 1. Example DYD1 on bulls in two countries

The state of the s				
Bull identification	Country of proof	Daughters	DYD	
		(no.)		
1	AA	10	56	
2	AA	20	-23	
3	AA	50	8	
1	BB	100	9	
4	BB	40	3	
5	BB	20	-11	

¹Daughter yield deviations.

Thus, heritabilities are .36 for AA and .235 for BB, and the genetic correlation is .89.

The pedigrees of all bulls in both countries are given in Table 2, and phantom parent groups are indicated as pedigrees P1 through P6. The inverse of the additive relationship matrix with phantom groups added is given in Tables 3 and 4 following the rules of Quaas (8), modified for computing relationships using sire, MGS, and maternal granddam (MGD) rather than sire and dam as per VanRaden et al. (13). The contributions to the inverse of the relationship matrix are as follows:

	Bull	Sire	MGS	MGD
Bull	х	5x	25x	25x
Sire	5x	.25x	.125x	.125x
MGS	25x	.125x	.0625x	.0625x
MGD	25x	.125x	.0625x	.0625x

where x = 16/(m + 11) and m = 0 if both sire and MGS are known (not phantom groups), m = 1 if the sire is known but MGS is unknown, m = 4 if the sire is unknown and the MGS is known, and m = 5 if both sire and MGS are unknown. The MGD is unknown in all cases and should be assigned to the appropriate phantom groups. These rules assume no inbreeding. VanRaden et al. (13) put all unknown sires and MGS into the same sire groups and put MGD into separate groups. For the multiple-trait model, separate phantom groups for sires, MGS, and MGD are recommended because of the different selection intensities that could be applied to each group.

There are dependencies among the group effect equations; to remove these dependencies, add 1 to the diagonals of the phantom group effects in the inverse of the relationship

TABLE 2. Pedigree information for example data.

	Ider				
Bull	Sire	MGS	MGD1	m ²	
1	6	7	P5	0	_
2	8	9	P5	0	
3	10	8	P5	0	
4	10	11	P6	0	
5	2	6	P6	0	
6	P1	P2	P6	5	
7	PΙ	P2	P6	5	
8	P1	P2	P6	5	
9	P3	P4	P6	5	
10	P3	P4	P6	5	
11	P3	P4	P6	5	

¹Maternal granddam.

 2 m = 0 if both sire and maternal grandsire (MGS) are known, = 1 if the MGS is unknown, = 4 if the sire is unknown, and = 5 if both sire and MGS are unknown.

matrix as discussed by Schaeffer (11). Then the group solutions will sum to 0. Therefore, the solutions (ETA) for bulls will be relative to the same genetic base within each country. The next subsection discusses the consequences of this approach.

The multiple-trait mixed model equations for this analysis need not be constructed explicitly but can be solved by iteration on the pedigree and data files (12). Solutions for bulls are expressed as ETA within each country (Table 5). Some countries express bull proofs as estimated breeding values, and so the results from this analysis would need to be doubled. The solutions for bulls are relative to the same genetic base for each country and have their respective functions of phantom group solutions already added into their solutions (8). To obtain the ETA of bull 1, for example, expressed with the genetic base currently used in country AA, the solution for $\mu_1 = 10.4966$ must be added to the bull's solution giving 10.4966 + 31.1325 = 41.6291. Similarly, the solution for $\mu_2 = 1.2141$ must be added to all bull solutions for country BB to be expressed in terms of the current genetic base used in country BB.

Fixed Versus Random Effects for Phantom Groups

A suggestion was made to add 1 to the diagonals of the phantom group effects for

TABLE 3. Upper left 11×11 matrix of inverse elements of additive genetic relationship matrix for pedigrees of example bulls.

ID	1	2	3	4	5	6	7	8	9	10	11
1	256	0	0	0	0	-128	-64	0	0	0	0
2	0	320	0	0	-128	32	0	-128	-64	0	0
3	0	0	256	0	0	0	0	64	0	-128	0
4	0	0	0	256	0	0	0	0	0	-128	-64
5	0	-128	0	0	256	-64	0	0	0	0	0
6	-128	32	0	0	-64	256	32	0	0	0	0
7	-64	0	0	0	0	32	192	0	0	0	0
8	0	-128	-64	0	0	0	0	256	32	32	0
9	0	64	0	0	0	0	0	32	192	0	0
10	0	0	-128	-128	0	0	0	32	0	304	32
11	0	0	0	-64	0	0	0	0	0	32	192

¹All values are multiplied by 176.

forming the inverse of the relationship matrix, i.e., to the diagonals of $\mathbf{Q'A^{-1}Q}$, where \mathbf{Q} is as described in the model, and $\mathbf{A^{-1}}$ is the inverse of the relationship matrix among bulls. This change implies that phantom group effects are random with expected values of zero. Refer to this approach as Groups Random. However, this approach could also be viewed as biased estimation of the fixed effects of phantom groups. That is, a small amount of bias of the estimates is accepted in exchange for the hope of getting smaller mean squared errors. The latter viewpoint is probably more correct be-

cause phantom group effects represent differences in the accumulated effects of previous selection and should not have expected values of 0.

If phantom group effects are fixed (i.e., ones are not added to the diagonals of $\mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q}$), then

$$\mathbf{Q}'\mathbf{A}^{-1}\hat{\mathbf{s}} = \mathbf{0},$$

and, if Groups Random, then

$$\mathbf{Q}'\mathbf{A}^{-1}\hat{\mathbf{s}} = \hat{\mathbf{g}}.$$

TABLE 4. Upper right 11×6 and lower right 6×6 matrix of inverse elements of additive genetic relationship matrix for pedigrees (P) of example bulls.¹

Identification	P1	P2	P3	P4	P5	P6
1	0	0	0	0	-64	0
2	0	0	0	0	-64	32
3	0	0	0	0	-64	0
4	0	0	0	0	0	-64
5	0	0	0	0	0	-64
6	88	-44	0	0	32	-28
7	-88	-44	0	0	16	-44
8	-88	-44	0	0	48	-44
9	0	0	-88	-44	16	-44
10	0	0	-88	-44	32	-12
11	0	0	-88	-4 4	0	-28
P1	308	66	0	0	0	66
P2	66	209	0	0	0	33
P3	0	0	308	66	0	66
P4	0	0	66	209	0	33
P5	0	Ō	0	0	224	0
P6	66	33	66	33	0	274

¹All values are multiplied by 176. Ones have been added to the diagonals of the phantom group effects.

2676 SCHAEFFER

TABLE 5. Solutions to multiple-country analysis.

Bull identification	Solution in country AA	Solution in country BB (BCA) ¹
	(kg)	
Mean Bulls	10.4966	1.2141
1	31.1325	7.0211
2	-26.5385	-5.9504
2 3	-2.4056	4772
4	4.3014	1.1245
5	-29.2208	-7.0674
6	10.9362	2.3169
7	8.9970	2.0117
8	-12.8806	-2.9245
9	-8.3029	-1.8438
10	1.4727	.4375
11	.0465	.0695
Phantom groups		
P1	2.5895	.5547
P2	1.2948	.2774
P3	9812	1526
P4	4906	0763
P5	1.5631	.3908
P6	-3.9756	9939

¹Breed class average.

These equalities are true within each country in the analysis. For the example data,

which is the same whether phantom group effects are treated as fixed or random. When groups are fixed, then the solutions for sires 6, 7, and 8 sum to 0, and the solutions for sires 9, 10, and 11 sum to 0. The last two lines impose other conditions on the sum of sire solutions that are difficult to interpret but that are due to the cross classification of MGD groups with sire and MGS groups and with bulls for which sire and MGS were known. With Groups Random, the solutions for phantom groups are functions of the bulls for which sires, MGS, and MGD belong to those groups. This function is identical to that for sires of bulls, i.e., their solution is a function of the solutions of their sons.

With large numbers of observations per phantom group, the ranking of elements within $Q\hat{g} + \hat{s}$ is nearly identical whether phantom groups are fixed or random. However, the bias (and standard errors) associated with small phantom groups could be large and might lead to differences in rankings of bulls in these groups compared with those for bulls in other groups.

The user of the multiple-trait approach should decide whether to have phantom group effects as either fixed or random. I prefer the Groups Random approach.

Conversion Formulas

With the multiple-trait analysis, each bull receives an ETA for every country in the analysis. Hence, a conversion formula is not needed to convert a bull's ETA from country AA to an ETA expressed in terms of country BB. However, for bulls not included in the analysis, a conversion formula may be needed. Conversion formulas may be derived following Wilmink et al. (16) or Goddard (3), or just the simple regression of one country's proofs on another country. The approaches of Wilmink et al. (16) and Goddard (3), as they are known, account for the accuracy or reliability of the ETA in each country. A comparison of these different approaches on actual data would be necessary.

Let PA and PB be the proofs of a bull in country AA and BB, respectively, equal to the sum of the sire solution and the solution for μ . Assuming simple regression is adequate, the regression of PA on PB gives an equation to predict PB from PA. Using the PA and PB on the 11 sires in the multiple-trait analysis, the conversion formula is

$$PB = -1.2058 + .22942(PA)$$
.

Similarly, the regression of PB on PA gives

$$PA = 5.2608 + 4.3518(PB).$$

Notice that rearranging the first prediction equation gives

$$.22942(PA) = PB + 1.2058$$

PA = 5.2557 + 4.3587(PB).

Thus, the two conversion formulas are almost completely reciprocal. Most conversion formulas available today are not reciprocal, which bothers many dairy producers. If the genetic correlations among countries were unity, then the conversion formulas should be exactly reciprocal. Perfect correlations would lead to a singular variance-covariance matrix of sire effects, and alternative computing strategies for singular G would be needed (6).

For computation of conversion formulas. INTERBULL has recommended use of only bulls with reliabilities >.749 and with daughters calving within the last 10 yr. With the multiple-trait analysis, all bulls could be used to derive the conversion formulas. Thousands of bulls could be in the multiple-trait analysis on which to base the conversion formulas, compared with the small numbers that are currently used to derive most conversion formulas. The conversion formulas derived from the multiple-trait sire ETA should also not be biased by preferential treatment because only the DYD from the country where they were first progeny tested are included in the analysis, i.e., the DYD, which may be subject to preferential treatment, of imported bulls are excluded.

DISCUSSION

The multiple-country method for international comparison described in this paper has been recommended for use by INTERBULL as a service to member countries. Estimation of genetic parameters within and between countries is a problem. The observations are weighted mean DYD. Because cows have different numbers of lactations, the appropriate residual variance and heritabilities to use with DYD are not clear. VanRaden and Wiggans (14) indicated that the variance of individual yield deviations is much less than for raw yields. The variance of a mean of yield deviations for a cow is proportional to [1 + (n -1)h²]/n, where n is the number of lactations going into the mean. Thus, the daughter means going into a bull's DYD have different variances for each daughter. An older bull would have many daughters with several lactations each, but a young sampling sire would have daughters with only first lactations. The DYD are similar to deregressed bull proofs and generally have a correlation with bull proofs of .99; the variance of DYD is only slightly greater than the variance of bull proofs (unpublished data). Thus, the ratio of residual to sire variances for DYD should not necessarily be equal to $(4 - h^2)/h^2$. Estimation of residual and sire variances by usual methods is not possible because the manner in which DYD are calculated does not allow the calculation of a total sum of squares of yield deviations for an individual cow.

The estimation of the genetic correlation between pairs of countries is equally difficult. If only the DYD of bulls in the country in which the bulls were first proven are used in the analysis, then crossproducts of bull ETA from the multiple-trait analysis will be highly dependent on the assumed genetic correlations. The number of sons of bulls that appear in other countries would also be important for obtaining good estimates. The problems of estimation of genetic parameters need a workable solution before the multiple-trait method can be used for international comparisons.

The multiple-country model should be better theoretically than that of Schaeffer (10). The multiple-country model accommodates different genetic parameters per country, different units of measurement per country, and genetic correlations between countries that are less than unity. If the genetic correlation between two countries is substantially less than 1. then sire ETA from the multiple-country analysis would not necessarily rank sires the same in those two countries. The success of any international comparison, however, is dependent on the amount and quality of connectedness of countries. Cross usage programs, using reference sires, may be needed to improve connectedness of certain countries so that their bulls may be ranked internationally.

Finally, the multiple-country model may be most applicable to conformation traits that are not measured or scored in the same manner in each country, and, therefore, the genetic correlations between countries may be low or negative. Further work is needed to apply these methods to production and conformation traits.

2678 SCHAEFFER

ACKNOWLEDGMENTS

The financial support of the Ontario Ministry of Agriculture and Food, the Canadian Association of Animal Breeders, Semex Canada, and Agriculture Canada is gratefully acknowledged. Thanks are expressed also for the help of Dale Van Vleck in revising this paper.

REFERENCES

- 1 Banos, G. 1992. Report on COPA/INTERBULL joint project. Int. Bull Eval. Serv. Bull. No. 7. Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.
- 2 Banos, G., L. R. Schaeffer, and E. B. Burnside. 1991. Genetic relationships and linear model comparisons between United States and Canadian Ayrshire and Jersey populations. J. Dairy Sci. 74:1060.
- 3 Goddard, M. 1985. A method of comparing sires evaluated in different countries. Livest. Prod. Sci. 13: 321
- 4 Henderson, C. R. 1977. Best linear unbiased prediction of breeding values not in the model for records. J. Dairy Sci. 60:783.
- 5 Henderson, C. R. 1984. Applications of linear models in animal breeding. Univ. Guelph, Guelph, ON, Canada.
- 6 International Bull Evaluation Service. 1990. Recommended procedures for international use of sire

- proofs. Int. Bull Eval. Serv. Bull. No. 4. Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.
- 7 Powell, R. L., and M. Sieber. 1992. Direct and indirect conversion of bull evaluations for yield traits between countries. J. Dairy Sci. 75:1138.
- 8 Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy Sci. 71:1338.
- 9 Rozzi, P., L. R. Schaeffer, E. B. Burnside, and W. Schlote. 1990. International evaluation of Holstein-Friesian dairy sires from three countries. Livest. Prod. Sci. 24:15.
- 10 Schaeffer, L. R. 1985. Model for international evaluation of dairy sires. Livest. Prod. Sci. 12:105.
- 11 Schaeffer, L. R. 1991. C. R. Henderson: contributions to predicting genetic merit. J. Dairy Sci. 74:4052.
- 12 Schaeffer, L. R., and B. W. Kennedy. 1986. Computing strategies for solving mixed model equations. J. Dairy Sci. 69:575.
- 13 VanRaden, P. M., E. L. Jensen, T. J. Lawlor, and D. A. Funk. 1994. Prediction of transmitting abilities for Holstein type traits. J. Dairy Sci. 73:191.
- 14 VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737.
- 15 Westell, R. A., R. L. Quaas, and L. D. Van Vleck. 1988. Genetic groups in an animal model. J. Dairy Sci. 71:1310.
- 16 Wilmink, J.B.M., A. Meijering, and B. Engel. 1986. Conversion of breeding values for foreign populations. Livest. Prod. Sci. 14:223.