International Dairy Bull Evaluations Expressed on National, Subglobal, and Global Scales

R. L. Powell and P. M. VanRaden
Animal Improvement Programs Laboratory, Agricultural Research Service,
United States Department of Agriculture,
Beltsville, MD 20705-2350

ABSTRACT

Genetic evaluations on a global scale were calculated for Holstein bulls using the May 2001 International Bull Evaluation Service (Interbull) evaluations expressed on each of 27 national scales. National scale data were weighted by the country's proportion of total daughters from all bulls (population size) to represent market share. Correlations between Interbull evaluations on national scales and evaluations on a global scale ranged from 0.961 to 0.998 (mean of 0.988). Number of top 100 bulls for protein yield that were in common between national and global scales ranged from 54 to 94 and was related significantly to mean genetic correlation between a country and the other 26 countries. Weighting of evaluations on national scales by population size, inverse of population size weight, or equal weight produced practically the same group of top bulls and correlations among the three global scales were 0.999. Thus, the method for combining Interbull evaluations expressed on national scales had only minor impact and was much less important than use of all data. Subglobal scales were established by a clustering technique that gave two to five groups. For grazing countries or other atypical systems, a subglobal scale may provide better guidance, although a scale representing three grazing countries did not provide the expected improvement over a global scale in the relationship with the three country scales. If conditions in nonparticipating countries are generally represented by participating countries, most needs are met by a global scale.
(**Key words:** genetic evaluation, international ranking, national scale, dairy cattle)

Abbreviation key: Interbull = International Bull Evaluation Service, **MACE** = multiple-trait, across-country evaluation, r_g = genetic correlation.

INTRODUCTION

The need for procedures to express genetic evaluations of dairy bulls across national borders has long been recognized. Meetings beginning in the early 1970s under the auspices of the European Association of Animal Production and the International Dairy Federation (Galliard et al., 1977) resulted in the formation of the International Bull Evaluation Service (**Interbull**) in 1983. Discussions at early Interbull meetings often dealt with proposed improvements to methods for calculating equations for conversion of evaluations from one national scale to another. Because conversion equations could be developed between only two countries at a time, calculation and application was tedious. Confusion and some mistrust developed between national dairy industries because the equations between two countries were computed by the importing country, which could be biased to serve local interest. Also, equations were not reciprocal, as some might have expected, because traits were not perfectly correlated across countries.

Schaeffer (1985) proposed an approach that combined national evaluations and used male pedigree information to provide ties between countries. Interbull implemented this procedure in August 1994 with an assumed genetic correlation between countries (r_g) of 1.0. Bull evaluations were presented on each national scale, but the ranking of bulls was the same in all countries. It was recognized, however, that rankings could and should differ on scales of different countries due to such factors as genotype-environment interaction and differences in data collection, evaluation systems, and trait definitions. Schaeffer (1994) provided a new method: multiple-trait, across-country evaluation (**MACE**), which uses estimates of r_g for each pair of countries. In February 1995, MACE was used for computing ease, but with an r_g of 0.995 for all country pairs because an estimation method for r_g had not been agreed upon (Interbull, 1995). Because $r_g \sim 1$, the ranking of bulls remained essentially the same for all national scales. In August 1995, estimated r_g were used for Interbull evaluations, resulting in unique ranking of bulls on each national scale.

Although rankings that varied by country were seen as appropriate, they created considerable challenges for nonparticipating countries. The assumption of $r_g = 1$ had allowed the use of any national scale for sire selection because bull rankings were the same in all countries. With $r_g \sim 1$, the ranking is different on the scale of each participating country so the country scale chosen by a nonparticipating country would impact on selection decisions. Although the goal is to use evaluations that best predict a bull's genetic merit in the local environment, the selection of which national scale to use could be based on many criteria including similarity of temperature, humidity, management techniques, evaluation system, and genetic background. The appropriate scale may not be obvious.

For international marketers, the existence of a trait evaluation on more than two dozen national scales makes representation of bulls difficult, particularly for sales efforts in nonparticipating countries or in international publications. An advertisement in Holstein International (2001) for example, included the protein ranking of a bull in nine different countries.

Selection on a global scale was suggested by van der Beek (1999). He investigated weighting Interbull evaluations by the largest eigenvector of the genetic correlation matrix. The result of this procedure was similar to a simple mean of national scales. A limited number of subglobal scales developed from scales of nations thought to have similar conditions could be more beneficial than a global scale or an individual national scale to nonparticipating countries and international marketers. The specific national scale would continue to be most appropriate for use within a participating country.

Weigel and Rekaya (2000) suggested clustering herds across national borders, essentially creating pseudo-countries with conditions more similar than exist in actual countries. In our study, clustering was of actual countries, which may be of more immediate use with current evaluation systems. Despite some similarities, the focus of the two approaches is different. Clustering herds is meant primarily to improve accuracy of evaluations. Our clustering is meant to provide genetic selection guidance for breeders in nonparticipating countries, breeders focusing on a global market, and to facilitate international marketing of semen.

The objective of this study was to examine methods for combining Interbull evaluations into a common global evaluation or subglobal evaluations and to compare the relationships between the developed scales and existing national scales. In this report, national scale means the results on a particular country scale as produced by Interbull, not the evaluations calculated by that country that were used as input to MACE.

MATERIALS AND METHODS

Data were May 2001 Interbull yield evaluations for 65,472 Holstein bulls born in 1984 or later (Interbull, 2001). Evaluations included information from nearly 24 million daughters with milk, fat, and protein records. Bulls were required to have a protein evaluation. Data were from 24 countries; however, red Holsteins are treated as separate populations in Denmark, France, and Switzerland. Thus, there were a total of 27 Holstein populations, and we will refer to 27 "country" scales.

Individual evaluations expressed on national scales were standardized by subtracting the national mean for the trait (milk, fat, or protein) from the bull evaluation, and then dividing by the national sire SD for the trait, provided by Interbull (Interbull, 2001). Nordic countries reported evaluations as relative breeding values, which were converted to kilograms as the Interbull SD were in kilograms.

Standardized evaluations on the national scales were combined and weighted by the country's population size to calculate global evaluations. Population size was the proportion of a given country's daughters of all bulls to the total daughters from all countries (Table 1), representing the market share among participating countries. To examine the impact of the weighting factor, global scales were also calculated using the reciprocals (inverses) of weights divided by sum of reciprocals, or using a simple mean of standardized evaluations (equal weighting). Population size weighting was used in other comparisons. The evaluations on all three scales were on an SD basis and could be rescaled if desired.

Correlations were computed between Interbull evaluations on each national scale as well as with global evaluations. The number of bulls in common in the top 100 bulls for protein on each national scale and the top 100 bulls on the global scale was determined, and the characteristics of countries with more or fewer bulls in common were examined.

February 1995 bull evaluations calculated with estimated r_g and unity r_g (actually 0.995) on national scales for Canada, Denmark, Finland, France, Germany, Italy, The Netherlands, Sweden, and the United States were available from a prior project (Powell and Norman, 2000). Global evaluations derived from the national scale evaluations calculated with estimated r_g were compared to the national evaluations calculated with unity r_g to indicate whether these two approaches might be essentially the same.

Subglobal scales were produced by combining subsets of national evaluations. Grouping of countries was based on cluster analysis by the VARCLUS procedure of SAS (SAS Institute, 1999). The procedure was applied to the correlation matrix of Interbull evaluations

Table 1. Proportion of all bull daughters by country, correlations between national and global scales, numbers of bulls in common between top 100 for protein on the national and global scales, and mean genetic correlation (r_g) with other countries.

Country	Proportion of daughters (%)	Correlation of national and global evaluation	Top bulls in common	Mean r_g
US	17.4	0.992	76	0.89
Germany	15.3	0.994	76	0.87
France	12.3	0.996	83	0.88
New Zealand	10.6	0.972	57	0.79
Netherlands	9.3	0.996	82	0.90
Canada	4.7	0.994	75	0.89
United Kingdom	4.7	0.992	80	0.89
Australia	4.5	0.968	56	0.80
Italy	4.3	0.992	79	0.87
Denmark	4.1	0.997	86	0.89
Poland	1.6	0.990	82	0.87
Sweden	1.5	0.994	82	0.88
Spain	1.2	0.996	92	0.88
Czech Republic	1.2	0.968	67	0.86
Belgium	1.2	0.997	86	0.89
Ireland	1.1	0.984	69	0.88
Hungary	1.0	0.994	80	0.87
Israel	0.9	0.992	80	0.86
Switzerland (Red)	0.8	0.996	83	0.88
Finland	0.8	0.994	80	0.88
Switzerland	0.4	0.993	78	0.88
South Africa	0.3	0.995	87	0.86
Estonia	0.3	0.995	91	0.86
Slovenia	0.1	0.998	91	0.86
Austria	0.1	0.991	85	0.86
Denmark (Red)	0.1	0.996	88	0.87
France (Red)	0.1	0.999	94	0.88

on each of the 27 national scales. Potential numbers of clusters ranged from two through five.

RESULTS AND DISCUSSION

Unless stated otherwise, results are for protein yield. Correlations of evaluations on national scales with the global scale (Table 1) ranged from 0.968 to 0.999 (mean of 0.991). This is higher than among national scales, which ranged from 0.915 to 0.995 (mean of 0.976). For all countries, the number of national top bulls in common with the global top 100 bulls was highly and significantly ($P < 0.001$) correlated with the correlation of national and global scales.

The number of top 100 bulls for protein yield that were in common between national and global scales ranged from 56 to 94 (mean of 80) and was significantly ($P < 0.001$) related to mean r_g between a country and all other countries. Not surprisingly, countries with the lowest mean r_g with other countries (Australia and New Zealand) had the fewest top bulls in common with the global scale. However, that relationship was not significant for the other 25 Holstein populations.

Estonia, France (Red), and Slovenia were among countries with the highest numbers (>90) of top bulls in common with the global scale. Their smaller populations (<1% of total daughters) might suggest difficulty

in estimation of r_g and, therefore, use of more subjective r_g (Interbull, 2001). Each of their r_g with the other countries tended to be similar. Use of the same r_g with all countries would essentially result in a global scale for foreign bulls. Israel, South Africa, and Switzerland (Red), also had <1% of total daughters and also estimation difficulties (Interbull, 2001) but had more modest numbers (80 to 87) of bulls in common with the global top 100. Although slightly more data were available for these countries, the amount of foreign bull use likely also plays a role. Without local daughters or granddaughters of top foreign bulls, there is limited opportunity for the global and national top lists to differ.

Variation in r_g with other countries was thought to be related inversely to number of top bulls in common. For instance, the United States had the highest variation in r_g, and its number of top bulls in common with the global top bulls (76) was below the mean for all countries of 80. However, the correlation between number of top bulls in common and variation in r_g with other countries was not significant.

A country with many of its own bulls among top bulls on its national scale (such as the United States and The Netherlands) will have fewer bulls on a global scale. Multiple regression to predict number of common top bulls from mean r_g, SD of r_g, and numbers of domestic bulls on national and global scales showed that only

Table 2. Correlations between global bull evaluations based on various weightings of International Bull Evaluation Service evaluations expressed on national scales (above diagonal) and numbers of bulls in common among top 100 bulls for protein yield (below diagonal).

Weighting method	Weighted by national number of daughters	Equal weighting	Inverse weighting
Weighted by population size		0.9997	0.9994
Equal weighting	97		0.9998
Inverse weighting	94	97	

effect of mean r_g was significant ($P < 0.05$), and that significance disappeared in a reanalysis without Australia and New Zealand.

Using inverse weighting to combine Interbull evaluations on national scales to a global ranking resulted in exactly the same mean number of top bulls in common between national and global scales as was found with the original weighting by population size. Mean correlation between Interbull evaluations on national scales and the inversely weighted global scale was 0.991. Applying equal weights to standardized Interbull evaluations on national scales did not result in important changes in top bulls. Correlations of Interbull evaluations on national scales with global evaluations based on equal country weighting were about the same as those with global evaluations weighted by population size and ranged from 0.968 to 0.999 (mean of 0.992). These results support those of van der Beek (1999).

Correlations were very high among the three global scales. Table 2 also shows the numbers of bulls in common among the top 100. Of the top 100 bulls on the global scales that were equally weighted or weighted by population size (directly or inversely), 97 were in common. The high correlations between global scales and the similarity of top bulls supports the conclusion of Powell and Norman (2000) and Weigel and Powell (2000) that the relevant issue is use of all data rather than how data are combined. Powell and Norman (2000) showed that use of estimated r_g does not produce more accurate MACE evaluations than if r_g is assumed to be ~1, which is one way to produce a global evaluation. Weigel and Powell (2000) demonstrated that the MACE approach to combining national evaluations was not more accurate than using conversion equations.

Bull evaluations from MACE with near unity r_g were similar to global evaluations calculated with data from the nine countries in 1995. With data with estimated r_g in the February 1995 data and weighted by population size, average correlation between national (estimated r_g) and global scales was 0.995 and between national ($r_g \sim 1$) and global scales was 0.998. These correlations show that the global scale derived from national scales is similar to the scale from $r_g \sim 1$, but not identical. Because the r_g among these nine countries were higher

than among the 27 countries evaluated in 2001, these global scale correlations are even higher. Average r_g among the nine countries was 0.91 compared with 0.87 among the 27 countries in 2001.

Table 3 presents the clusters of countries as produced by PROC VARCLUS from SAS. This procedure was applied to the correlation matrix of Interbull evaluations for protein, but clusters would likely be similar to that from use of the r_g as proposed by van der Beek (1999). The two-cluster groupings separated out the Czech Republic and the grazing countries of Australia, Ireland, and New Zealand as a second group. Austria, Great Britain, Poland, and Switzerland formed another group with three clusters. In the four-cluster result, the Czech Republic stood alone. With five-cluster analysis, Switzerland moved into a group of 11 other countries

Table 3. Clusters of countries from cluster analysis of Interbull protein evaluations.

	Number of clusters			
	2	3	4	5
Country	Cluster assignment of country			
Canada	1	1	1	1
Germany	1	1	1	1
Italy	1	1	1	1
Netherlands	1	1	1	1
United States	1	1	1	1
Hungary	1	1	1	1
Denmark (Red)	1	1	1	1
Finland	1	1	1	1
France	1	1	1	2
Belgium	1	1	1	2
Spain	1	1	1	2
Estonia	1	1	1	2
Switzerland (Red)	1	1	1	2
France (Red)	1	1	1	2
Slovenia	1	1	1	2
Israel	1	1	1	2
South Africa	1	1	1	2
Denmark	1	1	1	2
Sweden	1	1	1	2
Switzerland	1	2	2	2
Austria	1	2	2	3
Poland	1	2	2	3
Great Britain	1	2	2	3
Czech Republic	2	3	3	4
New Zealand	2	3	4	5
Australia	2	3	4	5
Ireland	2	3	4	5

separating from the largest group. These results for protein were not as easily explained as for milk where the four-cluster result grouped the grazing countries together, the Nordic countries together (less the red Holsteins in Denmark), and the Czech Republic and Poland together. A user of cluster evaluations, such as a farmer in a nonparticipating country, would be best served with the same cluster definitions applied to all yield traits. Further, as with herd clustering, it would be confusing if not chaotic for entities (herds or countries) to be part of different clusters for different traits or for successive runs. Logic and judgment would need to be applied to the results from cluster analysis and constraints applied.

A subglobal scale for grazing countries was calculated from evaluations from Australia, Ireland, and New Zealand. These countries had the lowest correlations with the global scale (Table 1). Correlations for Australia and New Zealand with the global scale were 0.968 and 0.972 but were 0.995 and 0.990, respectively, with the grazing subglobal scale. The correlation for Ireland differed little, going from 0.984 to 0.980. The numbers of top 100 bulls in common between national scales and grazing scale ranged from 38 to 54 for the other 24 countries and were 83 for Australia, 74 for New Zealand, and 58 for Ireland. Thus, although Ireland was in the cluster with Australia and New Zealand, its correlation and top 100 showed more similarity with the global scale than the grazing scale. Nearly all of the top 100 bulls for protein on the Australian (95) and New Zealand (95) scales were in the top 200 on the subglobal scale, but only 85 were in common for Ireland. For the other 24 countries, 59 to 77 bulls were in common. Equal weighting in calculating the subglobal scale improved the relationship between the grazing and Irish scales but still only 69 of the top 100 were in common. It appears that Ireland has conditions between those of Australia and New Zealand and those of the rest of the countries.

Clusters of countries were identified for two through five groups. Ireland was always in the cluster with Australia and New Zealand, which was logical because of the similarity of grazing, but the r_g of Ireland with Australia and New Zealand are only 0.83 and 0.82, respectively, lower than with any other countries. Because correlations of evaluations put Ireland in the cluster with Australia and New Zealand despite the low r_g, the correctness of the r_g is called into question.

Clustering identified grazing countries as a group, but other cluster characteristics were not so clearly defined for protein. Surprisingly, Irish evaluations were not more highly correlated with the grazing subglobal scale than with the global scale. Perhaps that is partly because Ireland had only about 7% of the daughters for

the three countries in the cluster and fairly low r_g with the other two countries.

CONCLUSIONS

Evaluations on a global scale were calculated by weighting of standardized Interbull evaluations on 27 national scales. Weighting of evaluations was fairly robust in that weighting by population size, inverse of those weights, or equal weights for each country produced practically the same group of top bulls. Correlations among the three global scales were 0.999. Thus, the method for combining Interbull evaluations on national scales had only minor impact.

The top 100 bulls globally differed substantially from the top bulls on a few national scales. Although a single global scale would support reasonable selection decisions in countries that do not participate in Interbull and be useful for multi-national marketing, defining groups of countries that have highly related evaluations and developing subglobal scales could also be useful. Then, nonparticipating countries could select an appropriate scale based on a cluster of countries with similar climate, feeding and management practices, and evaluation system. Cluster results would be more easily understood than the differences among many national scales. The problem being addressed is the number of scales that presently must be considered by international users. While some may argue that developing one or more scales adds to the problem, in fact, users of the global or subglobal scales could greatly reduce the array of genetic estimates they need to consider to those on a single scale.

One or more subglobal scales might be useful for the more extreme situations such as extensive grazing. However, the results with the grazing cluster were not uniform. Other similarities between countries such as evaluation system and genetic linkages would tend to place certain countries in a cluster. The clustering procedure and analyses of resulting subglobal scale are likely very dependent on the appropriateness of the r_g used in the Interbull evaluation process. As subglobal cluster results may differ across traits, logic would need to be applied to arrive at a definition of clusters for implementation. A global scale would likely be beneficial for most nonparticipating countries and in global advertising and would simplify sire selection with little loss of accuracy.

REFERENCES

Gaillard, C., J. Dommerholt, E. Fimland, L. Gol-Christensen, J. Lederer, A. E. McClintock, J. C. Mocquot, and J. Philipsson. 1977. A.I. bull evaluation standard for dairy and dual purpose breeds. Livest. Prod. Sci. 4:115–128.

Holstein International. 2001. Sersia France advertisement. Holstein Int. 8(6):inside front cover.

International Bull Evaluation Service. 1995. INTERBULL Centre routine genetic evaluation, February 1995. Uppsala, Sweden.

International Bull Evaluation Service. 2001. INTERBULL routine genetic evaluation for dairy production traits, May 2001. Online (http://www-interbull.slu.se/eval/may01.html). Accessed July 11, 2001.

Powell, R. L., and H. D. Norman. 2000. Impact of genetic correlations on accuracy of future evaluations. J. Dairy Sci. 83 (July). Online (http://www.adsa.org/jds/papers/2000/online/ jul1552.htm). Accessed Sept. 14, 2001.

SAS Online Doc, Version 8. 1999. SAS Inst., Inc., Cary, NC.

Schaeffer, L. R. 1985. Model for international evaluation of dairy sires. Livest. Prod. Sci. 12:105–115.

Schaeffer, L. R. 1994. Multiple-country comparison of dairy sires. J. Dairy Sci. 77:2671–2678.

Van der Beek, S. 1999. Exploring the (inverse of the) international genetic correlation matrix. Int. Bull Eval. Serv. Bull. No. 22:14B20. Dep. Anim. Breed. Genet., Uppsala, Sweden.

Weigel, K. A., and R. L. Powell. 2000. Retrospective analysis of the accuracy of conversion equations and multiple-trait, across-country evaluations of Holstein bulls used internationally. J. Dairy Sci. 83:1081–1088.

Weigel, K. A., and R. Rekaya. 2000. A multiple-trait herd cluster model for international dairy sire evaluation. J. Dairy Sci. 83:815–821.