Prediction of Genetic Correlations and International Breeding Values for Missing Traits

T. Mark,*[1] W. F. Fikse,† P. G. Sullivan,‡ and P. M. VanRaden§
*Department of Large Animal Sciences, Faculty of Life Sciences, University of Copenhagen, Grønnegårdsvej 8, 1870 C, Denmark
†Interbull Centre, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 75007 Uppsala, Sweden
‡Canadian Dairy Network, Guelph, Ontario, Canada N1G 4T2
§Animal Improvement Programs Laboratory (AIPL), Agricultural Research Service, USDA, Beltsville, MD 20705-2350

ABSTRACT

Prediction of genetic merit for missing traits is possible by combining available indicator traits. Indicator traits were combined using genetic correlations obtained from multiple regression equations of estimated genetic correlations among available indicator traits on variables explaining production circumstances and trait definitions. This prediction of missing traits was closer to actual breeding values than breeding values for any of the indicator traits. This was verified by evaluating clinical mastitis in each of the Nordic countries as a missing trait. The derived methodology was used to predict breeding values for clinical mastitis in the United States for local and international bulls with an average reliability of 43%.

Key words: genetic correlation, international genetic evaluation, udder health, prior information

INTRODUCTION

Direct information is often missing for a breeding goal trait due to difficulties in recording the trait. In this study, a trait is defined as missing if there is no systematic recording in the country of interest. Ignoring missing breeding goal traits in the selection of parents for the next generation results in suboptimal genetic progress. Hence, it is important to get as accurate predictions as possible for missing traits even though such predictions must be entirely based on indirect measures. Combinations of indirect measures may yield better predictions of genetic merit than any single indicator alone.

Genetic correlations ($\mathbf{r_G}$) between the missing trait and available indicator traits are required to obtain predictions for missing traits, but such r_G are usually not readily available. However, a missing trait may be available in a country other than the one of interest. If the missing trait is systematically recorded and evaluated in a foreign country, then r_G may be predicted from multiple regression models using various explanatory variables available in both the country where the trait is available and the concerned country where the trait is missing (Mark et al., 2006b). In some extreme cases with very weak genetic ties among available traits (e.g., Mark et al., 2005a), the problem of obtaining suitable r_G for missing traits may not be much different than obtaining suitable r_G among available traits.

The Interbull Centre applies a procedure to postprocess estimated r_G, which could be applied to obtain genetic correlations for missing traits as well. The rules associated with this procedure are largely based on expert intuition. However, applying similar structural models as found in Rekaya et al. (2001) and Mark et al. (2006b) to predict r_G seems more desirable as it allows simultaneous consideration of several explanatory effects and because it is less subjective.

Examples of missing traits are milk yield in China, fertility in Australia, and clinical mastitis (**CM**) in the United States. The latter will be the focus of this study, but the principles can be applied in other situations as well. Clinical mastitis is only recorded and used in genetic evaluations in the Nordic countries. Clinical mastitis information from these countries as well as milk somatic cell (**SC**; used herein to indicate both SCS and SCC) from the United States and other countries can be used as indirect measures (i.e., indicator traits) of CM in the United States.

International genetic evaluations provide an opportunity to incorporate CM information from Nordic countries into selection decisions in countries without direct CM information (Mark et al., 2002). However, current evaluations do not facilitate optimal use of the CM information in countries without CM records. This is because the CM information is converted to SC breeding values in such countries. More of the CM information

Table 1. Number of national breeding values (n), number of parities considered, average number of common bulls (CB),[1] average weighing factor (EDC), and heritabilities (h^2) per trait[2]

Trait[3]	Country	n	Parities	CB	EDC	h^2
CM	Denmark	4,338	3	124	491	0.04
	Finland	635	3	34	446	0.05
	Sweden	1,381	1	178	250	0.02
SC	Denmark	4,747	1	154	205	0.11
	Finland	656	3	37	351	0.17
	Sweden	1,402	1	178	213	0.08
	Canada	5,581	3	319	230	0.27
	Austria-Germany	12,207	3	347	342	0.23
	Estonia	273	3	16	316	0.12
	France	8,658	3	336	668	0.15
	United States	18,549	5	435	228	0.10

[1]Bulls having evaluations in each of 2 countries; CB among traits evaluated in the same country was excluded in this statistic.

[2]Data from Interbull February 2004 routine evaluation.

[3]CM = clinical mastitis, SC = milk somatic cell (SCC or SCS).

could be captured by directly relating CM in the Nordic countries, as well as SC in each country, with CM in the target country.

The aim of this study was to predict r_G for a missing trait, investigate the predictive performance of a method to predict international breeding values for missing traits, and determine the sensitivity of the method to the assumed r_G. This study focuses on applying the method to predict breeding values for CM in a country that has genetic evaluations for at least one correlated trait (i.e., SC).

MATERIALS AND METHODS

National genetic evaluation results for Holstein bulls for CM and SC from Denmark, Finland, and Sweden as well as SC from Canada, Austria-Germany, Estonia, France, and the United States were considered. The data were a subset of the data used in the February 2004 Interbull routine evaluation for udder health traits, and are described in more detail by Mark and Sullivan (2006). In summary, there were 4,338, 635, and 1,381 CM sire evaluations from Denmark, Finland, and Sweden, respectively. Furthermore, there were 52,073 sire evaluations for SC available for a total of 49,536 bulls with daughters in at least 1 of the 8 countries (Table 1). The heritabilities used in the national evaluations for clinical mastitis ranged from 0.02 to 0.05 whereas the heritabilities for SC ranged from 0.08 to 0.27. The number of parities considered for each trait and the average number of common bulls (**CB**) with evaluations in each of 2 countries also varied among the traits considered here (Table 1).

The REML (co)variance estimates among available traits were taken from Mark and Sullivan (2006). The r_G between SC in 2 different countries ranged from 0.80 to 0.96; the r_G between CM ranged from 0.71 to 0.86; and the r_G between SC and CM ranged from 0.51 to 0.73 within and across countries.

Variables Potentially Explaining Variation Among Genetic Correlations

Variables used to derive multiple regression equations for r_G in this study were obtained from 3 sources, and they could be grouped into 1) climatic variables, 2) production system indicators, and 3) national genetic evaluation descriptors.

The climatic variables were available from the Danish Meteorological Institute and were measured as the average monthly value during 1931 to 1960 in the capital city (Cappelen and Jensen, 2001). These averages were based on several daily measures. The daily minimum and maximum values were each averaged for every month, and the range was calculated as the difference between the highest average maximum and the lowest average minimum monthly value. The variables considered here were country averages of temperature (°C), range in temperature (from coldest to warmest month), country averages of rainfall (mm), range in rainfall, country averages of humidity (%), range in humidity, and country averages of wind speed (Beaufort scale). Squared terms of these variables were also considered.

Production system indicators were available from the International Committee for Animal Recording's yearly enquiries (ICAR, 2006). The most recent statistics were taken from each country. Holstein data were used when available; otherwise, statistics for all dairy breeds were used. The indicators considered were average milk yield (kg) and contents (%) of fat and protein from national milk recordings. Squared terms of these variables as

well as interactions among climatic variables and production system indicators were considered.

National genetic evaluation descriptors were taken from the forms that were available on Interbull's homepage (Interbull, 2004). The descriptors that were considered in this study were heritability, number of parities included, whether test-day records were considered, and whether the given trait was analyzed simultaneously with biologically different traits.

Finally, the CB and a variable explaining the effect of trait were considered. Trait was defined as a binary variable: trait = 1 if both involved traits were SC or if both involved traits were CM, and trait = 0 if one of the involved traits was SC whereas the other was CM.

Prediction of Genetic Correlations for Missing Traits

The estimated r_G were used as dependent variables in multiple linear regression to obtain regression coefficients that could be used to predict r_G involving missing traits. Explanatory variables in this regression were derived from the climatic variables, production system indicators, and national genetic evaluation descriptors described above.

The explanatory variables, except CB, were expressed as either ratios or binary variables. For continuous variables, a ratio was calculated so that the largest of the 2 country averages was in the denominator. Hence, $0 < \text{ratio} \leq 1$, and a high ratio always indicated that the variable in question was similar in the 2 countries. Likewise, a binary class variable was set equal to 1 if both traits belonged to the same class (e.g., both traits considered the same number of parities); otherwise, it was set equal to 0. The CB was used as is.

All variables were constructed so that the linear regression coefficient was expected to be positive. Variables with negative linear regression coefficient were dropped to ensure that the derived prediction formula would generalize well and be biologically meaningful when applied to missing traits. Effects with unexpected negative regression coefficients could be correlated with hidden confounders, which may take other values in the environment where the missing trait is expressed. The best model for r_G was selected based on Mallow's C(p).

Bending of Combined Genetic Correlation Matrix

The combined r_G matrix for both available and missing traits was not necessarily definitely positive. Therefore, the combined matrix was bent (Jorjani et al., 2004) before the prediction of breeding values for available and missing traits. In this weighted bending procedure, the diagonal elements of the r_G matrix were not allowed

to change, whereas the allowed changes for the genetic correlations were inversely proportional to the CB. The CB was arbitrarily set to 1,000 for correlations involving missing traits to allow only relatively small changes for the traits of main interest in this study. Changes in r_G due to bending were always ≤ 0.06.

Prediction of International Breeding Values for Available Traits

International breeding values for available traits were computed with a multiple-trait-multiple-country model (**MT-MACE**), which treats each country–trait combination (i) as a different but correlated trait (Schaeffer, 2001; Mark and Sullivan, 2006):

$$\mathbf{y}_i = \mu_i \mathbf{1} + \mathbf{Z}_i \mathbf{Q} \mathbf{g}_i + \mathbf{Z}_i \mathbf{s}_i + \mathbf{e}_i, \quad [1]$$

where $\mathbf{y}_i$ = vector of within-country univariately or multivariately deregressed national evaluations adjusted for residual correlations; μ_i = fixed effect of country-trait mean; $\mathbf{g}_i$ = vector of random genetic group effects; $\mathbf{s}_i$ = vector of random sire effects; $\mathbf{e}_i$ = vector of random residuals; $\mathbf{Z}_i$ = matrix assigning observations to sire effects; and $\mathbf{Q}$ = matrix assigning sires in $\mathbf{s}$ to group effects in $\mathbf{g}$. The (co)variance of the random variables was as follows:

$$\text{var} \begin{bmatrix} \mathbf{g}_i \\ \mathbf{s}_i \\ \mathbf{e}_i \end{bmatrix} = \begin{bmatrix} \mathbf{G}_0 \otimes \mathbf{I} & \mathbf{G}_0 \otimes (\mathbf{AQ}') & \mathbf{0} \\ & \mathbf{G}_0 \otimes \mathbf{A} & \mathbf{0} \\ \text{Symmetric} & & \mathbf{R}_i \end{bmatrix}$$

where $\mathbf{A}$ = the additive genetic relationship matrix relating bulls with their sires and maternal grandsires; $\mathbf{I}$ = an identity matrix; $\mathbf{G}_0$ = the genetic (co)variance matrix between traits; and $\mathbf{R}_i$ = the (co)variance among elements of $\mathbf{e}_i$; it is a diagonal matrix with diagonal elements equal to $\sigma^2_{e(i)}/\text{EDC}_{\text{MT}(i,k)}$ for bull k. The EDC_{MT} are effective independent weighting factors (Sullivan and Wilton, 2001; Mark and Sullivan, 2006) and $\sigma^2_{e(i)}$ are the residual variances. The residual variances are assumed equal to $(4\sigma^2_{\text{sire}(i)}/h^2_i) - \sigma^2_{\text{sire}(i)}$, where $\sigma^2_{\text{sire}(i)}$ is the sire variance and h^2_i is the heritability assumed in the national evaluations for each trait, respectively.

Prediction of International Breeding Values for Missing Traits

The vectors of MT-MACE solutions ($\mathbf{s}_i$) for each available country-trait (i) were subsequently combined into direct breeding values ($\mathbf{s}_{i+}$) for a missing trait (i+) using (Henderson, 1977):

$$\mathbf{s}_{i+}' = \mathbf{G}_{0ii+}' \, \mathbf{G}_{0ii}^{-1} \mathbf{s}_i', \qquad [2]$$

where $\mathbf{G}_{0ii+}$ = n × m matrix containing the expected genetic covariances between the m missing and n available traits, and $\mathbf{G}_{0ii}$ = n × n (co)variance matrix among the available traits. This formula is a generalization of the equation derived by Klei (1995) for a situation in which a bull has daughter information in only 1 country (i = 1) for a total of 2 countries:

$$\mathbf{s}_2 = r_g \frac{\sigma_{g_2}}{\sigma_{g_1}} \mathbf{s}_1,$$

where r_g = the genetic correlation between the available and missing trait, and σ_{g1} and σ_{g2} = the genetic standard deviation for the available and missing trait, respectively. All elements in $\mathbf{G}_{0ii}^{-1}$ and $\mathbf{s}_i$ are available when solving the MT-MACE equations, but $\mathbf{G}_{0ii+}$ needs to be specified. Note that the prediction formula is independent of reliabilities among breeding values for available traits so that the prediction does not need to be performed centrally.

Analyses and Comparisons

First, a MT-MACE analysis was conducted for all available traits. This analysis included SC from 8 countries and CM from 3 of these countries and was identical to the one presented by Mark and Sullivan (2006). The resulting breeding values from this analysis were labeled reference breeding values. Next, 3 analyses were performed to investigate the predictive performance of equation [2]. Here, either all the Danish, Finnish, or Swedish CM records, respectively, were set missing whereas the exact same (co)variance structure from the 11-trait reference evaluation was maintained. These analyses were repeated, but using predicted r_G based on prior information only. In each of these analyses, new prediction formulas for r_G were created by omitting estimated r_G involving the assumed missing trait. The reference breeding values were compared with the following 4 sets of breeding values from the analyses with a CM trait set missing:

1) breeding values for the trait treated as missing obtained using equation [2] and using estimated r_G for the actual trait;
2) breeding values for the trait treated as missing obtained using equation [2] and using predicted r_G;
3) breeding values for SC from the same country as the trait of interest; and
4) breeding values with the highest correlation to the reference breeding values.

Finally, 2 analyses involving 11 available traits and CM in the United States as a missing trait were conducted.

The potential loss of genetic progress (ΔG_{loss}) by using an alternative selection strategy was as follows:

$$\Delta G_{loss} = \frac{1}{100\sigma_{sire}} \left(\sum_{i=1}^{100} BV_i - \sum_{j=1}^{100} BV_j \right), \qquad [3]$$

where BV = the reference CM breeding values; σ_{sire} = the sire standard deviation in the reference evaluation; i = the ranking based on the reference breeding values; and j = the ranking based on breeding values for either direct, within-country SC or best correlated trait. All traits were standardized so that high breeding values were preferable.

Reliabilities were approximated using the information source method of Harris and Johnson (1998). Reliabilities for different groups of bulls were studied: 1) young bulls (i.e., bulls that were born in 1997 or later and had daughters in only one country). These were studied for both the domestic country (**d**) and the foreign (**f**) countries where the bulls have no daughters; 2) export bulls (i.e., bulls with daughters in at least 2 countries and most daughters in the given country); 3) import bulls (i.e., bulls with daughters in the given country, but most daughters in a country other than the given country). Thus, a single bull could be labeled as an export bull in only one country; at the same time being labeled an import bull in 1 to 7 countries.

RESULTS AND DISCUSSION

Model for Genetic Correlations

Many of the estimated effects for production system ratios were either negative or insignificant and therefore not considered in the final model for r_G. The following model for r_G, weighted by CB, explained 94% of the variance associated with estimated r_G and was preferred according to both Akaike's information criteria (Akaike, 1973) and Mallow's C(p):

$$r_G = \beta_0 + \beta_1(\text{trait}) + \beta_2(CB) + \beta_3(\text{parity}) + e. \qquad [4]$$

The estimates of β_0 ($P < 0.0001$), β_1 ($P < 0.0001$), β_2 ($P = 0.0007$), and β_3 ($P = 0.0024$) were 0.599, 0.275, 1.22×10^{-5}, and 0.0331, respectively. The estimated regression coefficients were sensitive to the omission of a clinical mastitis trait from the estimation (Table 2). However, the changes in regression coefficients partly counteracted each other, so the predicted r_G were robust. That is, the difference between r_G predicted from equation [4] and r_G predicted by a similar equation in

Table 2. Estimated regression coefficients ($\beta \pm$ SE) for equation [4] and similar estimates when clinical mastitis (CM) in Denmark, Finland, and Sweden, respectively, was set missing

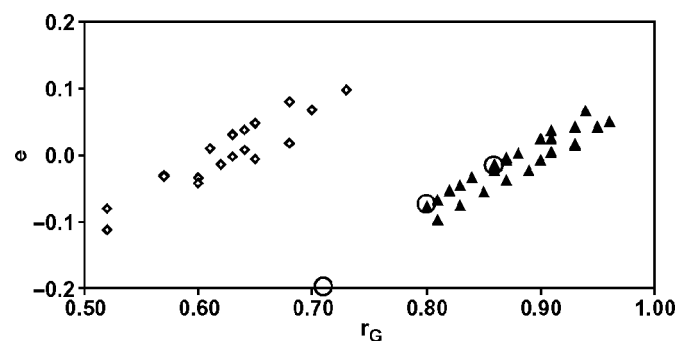| Missing trait	β_0	β_1	β_2 ($\times 10^{-5}$)	β_3
None	0.599 ± 0.013	0.275 ± 0.012	1.22 ± 0.34	0.0331 ± 0.0103
CM in Denmark	0.587 ± 0.014	0.279 ± 0.012	2.97 ± 0.77	0.0233 ± 0.0119
CM in Finland	0.604 ± 0.014	0.270 ± 0.012	1.11 ± 0.35	0.0384 ± 0.0104
CM in Sweden	0.583 ± 0.018	0.292 ± 0.015	1.46 ± 0.44	0.0278 ± 0.0135

which the concerned trait was not used to estimate the regression coefficients was always 0.02 or less, except for the within-country r_G between SC and CM in Denmark (difference = 0.07).

Analyses were repeated with different transformations of the dependent variable (r_G) to investigate if this would improve goodness of fit. Different power functions of r_G as well as logarithmic transformations were tested, but none gave a better fit than the untransformed r_G. This was in contrast to similar analyses carried out for milk yield (Mark et al., 2006b) in which r_G raised to the power of 5 gave the best fit. This may be because the r_G in the current study ranged from moderate to high, whereas all the r_G for milk yield were high and consequently the distribution of the dependent variable was skewed. The skewness was −0.80 and −0.21 for estimated r_G for milk and udder health, respectively. This illustrates that different models for r_G should be used for different traits or environments.

Figure 1 illustrates that the model for r_G regresses observations toward the average estimated r_G within each class of trait. This means that the predicted r_G involving missing traits, which are predicted based only on prior information, will vary less than estimated r_G among available traits. This seems desirable because no available trait is favored more than is supported by data in generating breeding values for the missing trait.

The largest absolute difference between estimated r_G and r_G predicted by formula was 0.197 for the r_G be-

tween CM in Finland and Denmark. This large difference is explained by a relatively low estimated r_G and that this estimate received relatively low weight in developing prediction equation [4] (only 35 CB). The estimated r_G between CM in Finland and Denmark was 0.71, whereas the weighted average of estimated across-country r_G for CM was 0.83 (weighted by CB), indicating that the r_G may be underestimated.

Estimated r_G, which are based on relatively few CB, could be severely underestimated (Sigurdsson et al., 1996; Mark et al., 2005a). Therefore, estimated r_G, which are based on few CB and which are lower than prior expectations, are regressed upward in routine international genetic evaluations performed by Interbull. A single large residual is therefore not necessarily an undesirable feature of prediction equation [4].

The average difference per trait between estimated and predicted r_G using prediction equation [4] ranged between −0.038 (CM in Finland) and 0.010 (SC in the United States); suggesting that there were no systematic bias in r_G for any trait (Table 3). The average bias of predicted r_G was also close to zero (i.e., −0.03) for within-country r_G between SC and CM.

Predicted r_G for CM in USA using equation [4] ranged between 0.874 and 0.878 with CM traits and ranged between 0.599 and 0.614 with SC measured in a different country. In comparison, the predicted r_G between

Table 3. Mean[1] estimated genetic correlation (r_G), mean error (ME),[2] and mean squared error (MSE)[2] of predicted genetic correlations

Trait[3]	Country	r_G	ME	MSE
CM	Denmark	0.65	−0.030	0.0060
	Finland	0.67	−0.005	0.0062
	Sweden	0.66	−0.006	0.0021
SC	Denmark	0.80	−0.007	0.0018
	Finland	0.77	−0.038	0.0043
	Sweden	0.81	0.005	0.0010
	Canada	0.81	−0.006	0.0020
	Austria-Germany	0.80	−0.011	0.0034
	Estonia	0.81	−0.003	0.0018
	France	0.82	0.002	0.0003
	United States	0.81	0.010	0.0016

[1]Summarized per trait (average of n = 10 correlations).
[2]$ME = \sum_{i=1}^{n}(r_{G_i} - \hat{r}_{G_i})/n, MSE = \sum_{i=1}^{n}(r_{G_i} - \hat{r}_{G_i})^2/n, \hat{r}_G$ = predicted genetic correlation using equation [4]; r_G = REML estimate.
[3]CM = clinical mastitis, SC = milk somatic cell (SCC or SCS).

Figure 1. Residuals (e) from prediction of r_G using equation [4] as a function of estimated r_G for r_G between clinical mastitis (CM) and milk somatic cell (SC) ($\diamond$), between CM and CM ($\circ$), and between SC and SC ($\blacktriangle$), respectively.

Table 4. Predicted genetic correlations between clinical mastitis (CM) in the United States and various nonmissing indicator traits obtained without (unforced)[1] or with forced[2] harmonization of arbitrary differences between traits

Trait[3]	Country	Unforced	Forced
CM	Denmark	0.876	0.909
	Finland	0.874	0.907
	Sweden	0.878	0.911
SC	Denmark	0.601	0.634
	Finland	0.599	0.632
	Sweden	0.603	0.636
	Canada	0.614	0.647
	Austria-Germany	0.610	0.643
	Estonia	0.599	0.632
	France	0.611	0.644
	United States	0.647	0.647

[1]Unforced genetic correlations obtained using equation [4].

[2]Forced genetic correlations obtained using equation [4] in which parity was set equal to 1.

[3]CM = clinical mastitis, SC = milk somatic cell (SCC or SCS).

CM and SC within the United States was 0.647 (Table 4).

Reasons for genetic correlations among countries being less than unity include genotype × environment interaction as well as differences in trait definition and genetic evaluation systems. It may be desirable in some situations to eliminate sources of variation not fully related to genotype by environment interaction when predicting r_G for missing traits. The effect of forced harmonization of trait definitions was illustrated with the variable parity, although perhaps not desirable in this case. The variable parity was forced to be identical for both CM in the United States and all available traits (Table 4). All the predicted r_G between CM in the United States and the available traits increased by 0.0331 when forcing the variable parity to be 1, except the United States within-country r_G because the same number of parities were already assumed within country (Table 4). If there is no obvious reason to consider parity 3 rather than parity 1 in the breeding goal, the arbitrary difference could be eliminated with respect to missing traits. Otherwise, no forced harmonization should be performed.

In a previous study, Mark et al. (2006a) used the estimated r_G for SC between 2 countries to determine the r_G for CM between the same 2 countries. In the current study, another strategy was followed for predicting r_G involving missing traits because the r_G for CM should not necessarily follow the r_G for SC measured in the same 2 countries. This could, for instance, be the case when one country only considers first-parity information for SC, but 3 parities for CM, which was the case for Denmark here. In addition, some definitions of SC may better describe CM than others; for example, when test-day information is used in certain ways.

The present approach tried to simultaneously utilize all the different similarities between traits (i.e., in terms of their definitions, their genetic evaluation model characteristics, and the environmental conditions in which they are expressed) that gives rise to high estimated r_G to predict r_G for missing traits. The variables included in prediction equation [4] were therefore based on identified reasons for variation in estimated r_G.

The model that was preferred in this study was rather simple because observations did not allow more detailed modeling for effects such as r_G type. Only 2 types of r_G were considered here: 1) across-country r_G between SC and SC as well as across-country r_G between CM and CM measured in different countries, and 2) within- and across-country r_G between SC and CM. Each of these groups was initially split into 2 groups separating r_G for SC from r_G for CM and separating r_G between SC and CM within and across countries. However, across-country r_G between SC and SC was not significantly ($P = 0.18$) different from across-country r_G between CM and CM, which can be explained by the fact that essentially only one reliable across-country estimate of r_G between CM and CM was available (i.e., the r_G between CM in Denmark and Sweden, because r_G involving Finland was based on low CB). Also, the effect of a binary class variable to distinguish between pairs of traits measured in the same or different countries was not significant ($P = 0.17$), which is likely because the effect of CB already explains some variation due to this.

The approach taken here to predict r_G has the advantage that it may be used when there are no indicator traits measured in a certain environment, provided that the environment in question does not deviate greatly from the environments in which the correlated traits were measured. The prediction formula, which was estimated in the current study, should probably not be used for environments that differ noticeably from the environments considered here. Torsell (2007) used the prediction formula of Mark and Sullivan (2006) to predict r_G for milk yield in Argentina. Although they found no difference between the average predicted r_G and estimated r_G, the correlation between predicted r_G and estimated r_G was almost zero. This illustrates that care should be given to extrapolation properties of equations to predict r_G for countries with deviating production circumstances.

The countries considered in this study did not vary much in terms of climate and production system indicators. This could explain why these variables were not important to include in the final model for r_G. In addition, the capital city may not represent average production circumstances well. If knowledge of the distribution of cows within countries were available, climate

Table 5. Correlation between reference and clinical mastitis breeding values obtained by direct prediction (direct); correlation between reference and within-country SC breeding values; and best correlation between reference and correlated breeding values for clinical mastitis in Denmark, Finland, and Sweden for either domestic,[1] foreign,[2] or all bulls

Item	Domestic bulls			Foreign bulls			All bulls		
	Denmark	Finland	Sweden	Denmark	Finland	Sweden	Denmark	Finland	Sweden
Direct (estimated r_G)	0.63	0.60	0.82	0.91	0.99	0.98	0.87	0.98	0.98
Direct (predicted r_G)	0.65	0.55	0.80	0.91	0.93	0.96	0.87	0.93	0.96
Within-country SC[3]	0.62	0.61	0.76	0.85	0.88	0.88	0.82	0.88	0.88
Best correlated trait	0.63	0.62	0.81	0.90	0.95	0.95	0.87	0.94	0.95

[1]Bulls with most daughters in the given country.

[2]Bulls with most daughters in country other than the given.

[3]SC = milk somatic cell (SCC or SCS).

conditions in dense cattle areas could be given more weight.

Today, international genetic evaluations for udder health also include data from warm countries such as Australia, South Africa, and Spain as well as from countries with year-round grazing such as Ireland and New Zealand. The best model to predict r_G would probably include additional explanatory effects if data from these countries were considered. Including more available traits in developing the best model for r_G would be beneficial as it could increase the robustness of the prediction formula to new environments and because the number of observations (i.e., estimated r_G) increases nearly quadratically as a function of the number of available traits considered.

Usefulness of Breeding Values for Missing Traits

Breeding values obtained with equation [2] for assumed missing traits were closer to reference CM breeding values compared with SC breeding values for the same country and with CM breeding values for a different country (Table 5). This was especially the case for export bulls. The use of predicted r_G reduced the correlation between reference breeding values and breeding values for the assumed missing trait, except for CM in Denmark.

It was expected that using the same r_G as in the reference evaluation would yield the highest correlation between breeding values. The different observations for CM in Denmark for domestic bulls can be explained by the within-country r_G between CM and SC, which increased from 0.51 in the reference evaluation to 0.73 in the evaluation assuming CM in Denmark to be a missing trait. This meant that, in the reference analysis, the r_G with CM in Denmark was greater for SC measured in countries other than Denmark. When SC in Denmark became relatively more important for predicting CM in Denmark compared with SC in other countries, the correlation between direct breeding val-

ues for the missing trait and the reference breeding values also increased. In most cases, it would seem logical that traits measured in the same country are more correlated than if they were measured in different countries. However, differences between traits within country in parities considered can decrease the within-country r_G. The relatively low r_G of 0.51 was lower than the estimate based on the international data (0.65) because it was forced to be equal to the r_G used in the Danish multiple-trait national genetic evaluation for udder health (Mark and Sullivan, 2006).

Within-country SC is not necessarily the best alternative to direct breeding values for missing traits (Table 5). For example, the best-correlated trait was CM in Sweden when the trait of interest was CM in Denmark. Similarly, SC in Germany-Austria and CM in Denmark had the highest correlation with reference breeding values for CM in Finland and Sweden, respectively.

The choice of selection strategy for the missing trait had a noticeable effect on which bulls had the best breeding values and on the potential genetic progress that could be achieved (Table 6). The potential loss of genetic progress from selecting the bulls with the 100 highest breeding values was lower for direct CM breeding values compared with breeding values for any other trait than the given. This was the case when either estimated or predicted r_G were used in the international evaluation, although the superiority of selecting for the direct trait was less clear with predicted r_G.

The average reliability of CM breeding values for the United States was 36.7 and 43.7% for young domestic and young foreign bulls, respectively (Table 7). The reason why the reliability was higher for bulls with daughters in countries other than the United States is partly because the coheritability for SC measured in the United States was lower than for most SC traits measured in foreign countries. For example, the coheritability with CM in the United States for SC in Canada was $0.27 \times 0.614 = 0.166$, whereas for SC in the United States, it was $0.10 \times 0.647 = 0.065$. Another reason for

Table 6. Potential loss of genetic progress (ΔG_{loss}) and percentage of coselected bulls in top 100 rankings in Denmark, Finland, and Sweden when using either clinical mastitis breeding values obtained by direct prediction (direct), within-country SC,[1] or best correlated trait compared with top 100 reference breeding values

	ΔG_{loss}			Percentage coselected		
Item	Denmark	Finland	Sweden	Denmark	Finland	Sweden
Direct (estimated r_G)	0.403	0.025	0.041	18	79	79
Direct (predicted r_G)	0.359	0.148	0.068	29	53	72
Within-country SC	0.530	0.283	0.256	19	37	41
Best correlated trait	0.427	0.134	0.091	19	49	65

[1]SC = milk somatic cell (SCC or SCS).

the relatively low average reliability for domestic bulls in the United States was that the weighting factor for SC was, on average, substantially lower in the United States than in the countries with second and third most records in the analysis (i.e., Austria-Germany and France; Table 1). The fact that there are several levels of approximations in computing reliabilities for international breeding values (e.g., weighting factors at the national level and the Harris-Johnson procedure at the international level) may also contribute to the differences between average reliabilities.

The average reliability of CM breeding values for Denmark decreased from 69.3 to 31.7% for young domestic bulls when Danish CM data were excluded (Table 7). The average reliability of young foreign bulls also decreased because parent averages were less accurate when direct data were excluded. This illustrates the benefit of having direct data for a breeding goal trait.

The relatively low correlations between reference and alternative breeding values for bulls with most daughters in the given country (Table 5) also show that there

was no substitute for considering data for the trait of interest in the international genetic evaluation, even though breeding values for missing traits were useful. This was especially the case when the domestic bulls were assumed competitive with the best foreign bulls. There were mostly foreign bulls in the top 100 ranking for CM in Finland and Sweden. Therefore, the potential loss of genetic progress (Table 6) was smaller for Finland and Sweden than for Denmark.

Average reliabilities increased when parity was forced to be equal in prediction equation [4] for r_G (Table 7). However, reliabilities were approximated assuming that genetic parameters were known without uncertainty. Accounting for uncertainty of genetic parameters would result in lower reliabilities (Mark et al., 2005b).

CONCLUSIONS

A method to predict genetic correlations and breeding values for missing traits was applied to udder health

Table 7. Average reliability (%) of breeding values for selected traits[1] and groups of bulls

	Missing traits			Available traits		
Bulls[2]	US CM[3]	US CM[4]	Denmark CM	US SC	Denmark CM	Denmark SC
Number of reliabilities						
Young (d)	1,672	1,672	764	1,672	811	811
Young (f)	3,020	3,020	8,591	3,020	8,591	8,591
Import	407	407	123	407	123	176
Export	284	284	47	284	47	47
All	51,490	51,490	51,473	51,490	51,490	51,490
Average reliability						
Young (d)	36.7	37.7	31.7	63.3	69.3	72.1
Young (f)	43.7	46.8	38.7	65.1	41.4	57.3
Import	47.7	50.5	59.4	84.2	84.4	87.5
Export	52.6	54.8	50.9	93.1	91.1	92.8
All	43.4	45.8	40.1	68.7	46.1	62.4

[1]CM = clinical mastitis, SC = milk somatic cell (SCC or SCS).

[2]Bull groups: d = domestic, f = foreign.

[3]Unforced genetic correlations obtained using equation [4].

[4]Forced genetic correlations obtained using equation [4] in which the variable parity was set equal to 1 for r_G involving missing traits.

data from several countries. The equation to predict r_G explained 94% of the variation among estimated r_G. Use of these predicted r_G yielded breeding values that may enable more efficient selection for resistance to CM than SC in countries without systematic recording of CM. The method may also be used to predict breeding values for countries that do not participate with any data in current international genetic evaluations provided that the production system does not deviate noticeably from the production systems for which information is available. Although the direct prediction of missing traits was useful, reliability was always higher when data of the trait of interest were included in the international genetic evaluation.

REFERENCES

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267–281 in Proc. Second Int. Symp. Informational Theory. B. N. Petrov and F. Csaki, ed. Akademiai Kiado, Budapest, Hungary.

Cappelen, J., and J. J. Jensen. 2001. Jordens Klima – Guide til vejr og klima i 156 lande. Teknisk rapport 01-17. Danmarks Meteorologiske Institut, Copenhagen, Denmark. (in Danish)

Harris, B., and D. Johnson. 1998. Information source reliability method applied to MACE. Interbull Bull. 17:31–36.

Henderson, C. R. 1977. Best linear unbiased prediction of breeding values not in the model for records. J. Dairy Sci. 60:783–787.

ICAR. 2006. Yearly enquiries. www.icar.org Accessed October 4, 2006.

Interbull. 2004. Description of national GES as applied in member countries. www.interbull.org Accessed October 4, 2006. Interbull, Uppsala, Sweden.

Jorjani, H., L. Klei, and U. Emanuelson. 2004. A simple method for weighted bending of genetic (co)variance matrices. J. Dairy Sci. 86:677–679.

Klei, L. 1995. Evaluating Holstein sires for conformation traits using data from the United States and the Netherlands. PhD Thesis. Cornell University, Ithaca, NY.

Mark, T., W. F. Fikse, U. Emanuelson, and J. Philipsson. 2002. International genetic evaluations of Holstein sires for milk somatic cell and clinical mastitis. J. Dairy Sci. 85:2384–2392.

Mark, T., W. F. Fikse, P. G. Sullivan, and P. M. VanRaden. 2006a. Prediction of international breeding values for non-measured traits: Application to clinical mastitis. Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil.

Mark, T., P. Madsen, J. Jensen, and W. F. Fikse. 2005a. Short communication: Difficulties in estimating across country genetic correlations for weakly linked bull populations. J. Dairy Sci. 88:3303–3305.

Mark, T., P. Madsen, J. Jensen, and W. F. Fikse. 2005b. Prior (co)variances can improve multiple-trait across-country evaluations of weakly linked bull populations. J. Dairy Sci. 88:3290–3302.

Mark, T., and P. G. Sullivan. 2006. Multiple-trait-multiple-country genetic evaluations for udder health traits. J. Dairy Sci. 89:4874–4885.

Mark, T., P. G. Sullivan, W. F. Fikse, and P. M. VanRaden. 2006b. Prior genetic correlations and non-measured traits. Interbull Bull. 35:72–75.

Rekaya, R., K. A. Weigel, and D. Gianola. 2001. Application of a structural model for genetic covariances in international dairy sire evaluations. J. Dairy Sci. 84:1525–1530.

Schaeffer, L. R. 2001. Multiple trait international bull comparisons. Livest. Prod. Sci. 69:145–153.

Sigurdsson, A., G. Banos, and J. Philipsson. 1996. Estimation of genetic (co)variance components for international evaluation of dairy bulls. Acta Agric. Scand. A, Anim. Sci. 46:129–136.

Sullivan, P. G., and J. W. Wilton. 2001. Multiple-trait MACE with a variable number of traits per country. Interbull Bull. 27:68–72.

Torsell, A. 2007. Prospects of performing multiple-country comparison of dairy sires among countries not participating in Interbull international breeding evaluations. MS Thesis no. 293. Swedish University of Agricultural Sciences, Uppsala.