Original article

Use of covariances between predicted breeding values to assess the genetic correlation between expressions of a trait in 2 environments

DR Notter, C Diaz

Virginia Polytechnic Institute and State University,
Department of Animal Science, Blacksburg, VA 24061, USA

Summary – Procedures to interpret correlation and regression coefficients involving predicted breeding values (BV) calculated for the same animals in different environments have been developed. Observed correlations are a function of the additive genetic correlation between performances in the 2 environments but are also affected by selection of animals that produce data in both environments, the accuracy of BV predictions in each environment, relationships among animals within and across environments and covariances among BV predictions within an environment arising from estimation of fixed effects in best linear unbiased prediction (BLUP) of animal BV. Methods to account for effects of selection and variable accuracy and experimental designs to minimize effects of relationships and covariances among BV predictions from estimation of fixed effects have been described. The regression of predicted BV in environment 2 on predicted BV in environment 1 is generally not affected by selection in environment 1, but both correlation and regression coefficients are sensitive to covariances among breeding value predictions within environments. In general, caution must be exercised in interpreting observed associations between predicted breeding values in different environments.

predicted breeding value / genetic correlation / regression / selection index / best linear unbiased prediction

Résumé – Utilisation des covariances entre les valeurs génétiques prédites pour estimer la corrélation génétique entre les expressions d'un caractère dans 2 milieux. *Des procédures sont établies pour interpréter les coefficients de corrélation et de régression impliquant des valeurs génétiques prédites (VG) calculées pour les mêmes animaux dans différents milieux. Les corrélations observées sont fonction de la corrélation génétique entre les performances dans les 2 milieux, mais elles dépendent aussi de la sélection des animaux sur lesquels des données sont recueillies dans les 2 milieux, de la précision des prédictions de VG dans chaque milieu, des parentés entre animaux intra-milieu et entre milieux et des covariances entre les prédictions de VG dans un milieu qui résultent de l'estimation des effets fixés dans la meilleure prédiction linéaire sans biais (BLUP) de*

VG. L'article présente des méthodes pour prendre en compte les effets de la sélection et de la précision variable des prédictions et des plans d'expérience pour minimiser les effets de la parenté et des covariances entre les prédictions de VG à partir de l'estimation des effets fixés. La régression des VG prédites dans le milieu 2 en fonction des VG prédites dans le milieu 1 n'est généralement pas affectée par la sélection dans le milieu 1, mais la corrélation et la régression sont toutes deux influencées par les covariances entre les prédictions de VG intra-milieu. D'une manière générale, une grande prudence est requise dans l'interprétation d'associations entre des valeurs génétiques prédites dans différents milieux.

prédiction de valeur génétique / corrélation génétique / régression / indice de sélection / meilleure prédiction linéaire sans biais

INTRODUCTION

Procedures to estimate additive genetic correlation (r_G) between expressions of the same trait in different environments were introduced by Falconer (1952), Robertson (1959), Dickerson (1962) and Yamada (1962). The procedures are analogoes to those for estimation of genetic correlation between 2 traits in the same environment, but recognize that performance is normally not measured on the same animal in multiple environments. Instead, related animals (often half-sibs) are produced in each environment and r_G is derived by comparing the resemblance among relatives in different environments to that observed among relatives in the same environment.

In single-generation experiments utilizing half-sibs, sires can produce progeny in pairs of environments. If sires are evaluated in environment 1 before being used in environment 2, divergent selection of sires can increase precision of estimates of the genetic regression of one trait on the other when a fixed number of progeny is measured (Hill, 1970; Hill and Thompson, 1977). This strategy makes use of the fact that selection of sires biases correlation between parent predicted breeding value (BV) and offspring performance but does not affect the regression of progeny performance on parent predicted BV so long as there is no selection of progeny records.

Data from industry performance-recording programs often include records of relatives evaluated in different environments, but the data structure is not under experimental control. Animals differ in the amount of information available, and unknown non-genetic sources of resemblance among relatives can exist. Likewise, little information may exist on procedures used to select parents in each environment. Procedures to estimate additive genetic covariances from these industry data sets exist (Meyer, 1991) and have been used to estimate covariances between expressions of the same trait in different environments (*eg*, Dijkstra *et al*, 1990). These analyses require that the model includes all genetic and nongenetic sources of resemblance among relatives and that information used to select parents be included in the data. Large numbers of records often exist, but only a fraction of them may represent records of relatives in different environments. Restriction of data to only records of animals with close pedigree ties across environments is tempting to reduce computational requirements, but may violate assumptions regarding selection.

If predicted BV for the same animals in different environments were derived using only data from within each environment, correlations among predicted BV across environments should provide information about r_G. Observed correlations in such situations (Oldenbroek and Meijering, 1986; DeNise and Ray, 1987; Tilsch *et al*, 1989a,b; Mahrt *et al*, 1990) were usually < 1, but, as noted by Calo *et al* (1973) and Blanchard *et al* (1983), the expected value of the correlations is also < 1, even if the underlying genetic correlation is unity. Thus correlations between predicted BV in different environments must be interpreted relative to their expected value.

This paper will consider the expected values of observed correlations and consider alternative experimental designs. Expected values of correlation and regression coefficients under ideal conditions will first be reviewed. Effects of non-random selection, variation in the accuracy of BV predictions, relationships among animals within and across environments, and covariances among BV predictions arising from estimation of fixed effects under best linear unbiased prediction (BLUP) will then each be considered.

RELATIONSHIP BETWEEN PREDICTED BV IN 2 ENVIRONMENTS

Let a population in environment 1 have additive genetic variance $\sigma_{u_1}^2$ for some trait. Predict BV ($\widehat{u}_1$) in that environment, and choose m sires to produce progeny in environment 2. Predict BV in environment 2 ($\widehat{u}_2$) using only data from that environment. Let the additive genetic variance for the trait in environment 2 be $\sigma_{u_2}^2$ and the genetic correlation between BV in environment 1 (u_1) and 2 (u_2) be r_G. Let the accuracy of BV prediction, a_1 and a_2 for environments 1 and 2, respectively, be the correlation between actual and predicted BV and be constant within each environment.

Under certain conditions, the expected correlation between predicted BV in the 2 environments ($r_{\widehat{u}_1\widehat{u}_2}$) is $a_1 r_G a_2$ and r_G can be estimated as $r_G = r_{\widehat{u}_1\widehat{u}_2}/a_1 a_2$. The conditions include (Taylor, 1983): 1) no environmental correlation between performance in the different environments; 2) no relationships among parents of measured animals; and 3) no other covariances among predicted BV within either environment. An additional assumption (4) is that sires are chosen at random. For sire evaluation with these assumptions, $a_{ij}^2 = n_{ij}/(n_{ij} + \lambda)$ where n_{ij} is the number of progeny for sire i in environment j and λ is the ratio of residual to sire variance. Assumption (1) is normally met if different animals are measured in different environments. Assumptions (2) and (4) can be met through choice of sires. Assumption (3) will not normally hold for BLUP, but may approximately hold under some conditions.

The regression (b) of $\widehat{u}_2$ on $\widehat{u}_1$ has the expectation:

$$b_{\widehat{u}_2\widehat{u}_1} = a_2^2 r_G \sigma_{u_2}/\sigma_{u_1}$$

such that $r_G = b_{\widehat{u}_2\widehat{u}_1}(\sigma_{u_1}/a_2^2\sigma_{u_2})$. Taylor's (1983) assumptions are required for this expectation, but random selection of sires is not. Knowledge of $\sigma_{u_1}^2$ and $\sigma_{u_2}^2$ is required to calculate the expected regression coefficient and for prediction of u_1 and

u_2. Effects of using incorrect values of σ_u^2 on estimates of r_G will not be considered further, but may be important.

Expected confidence limits for observed correlation $(\widehat{r})$ and regression coefficients $(\widehat{b})$ can be used to evaluate experimental designs in terms of their ability to detect significant departures of $\widehat{r}$ and $\widehat{b}$ from their expectations. For correlation analysis, Fischer's $z = 0.5[\ln(1+r) - \ln(1-r)]$ (Snedecor and Cochran, 1967) has variance of $\approx (m-3)^{-1}$ where m is the number of sires in the sample. For $|r| \leqslant 0.65$, confidence bounds on $\widehat{r}$ are of similar width at fixed m, whereas for $|r| > 0.65$, confidence bounds narrow with increasing r. Large numbers of sires are thus required if accuracies are low to avoid confidence limits that overlap 0.

For regression analysis, variance of $\widehat{b}$ $[V(\widehat{b})]$ is:

$$V(\widehat{b}) = \sigma^2_{u_2|\widehat{u_1}} / SS(\widehat{u}_1)$$

(Snedecor and Cochran, 1967) where $\sigma^2_{u_2|\widehat{u_1}}$ is the variance in $\widehat{u}_2$ at a fixed value of $\widehat{u}_1$ (*ie*, the mean square for deviations from regression) and $SS(\widehat{u}_1)$ is the sum of squares for $\widehat{u}_1$. Given Taylor's (1983) assumptions for m sires sampled at random from environment 1, $V(\widehat{b})$ is:

$$V(\widehat{b}) = [(1 - a_1^2 a_2^2 r_G)a_2^2\sigma^2_{u_2}]/(m-1)a_1^2\sigma^2_{u_1}.$$

Numbers of sires and progeny required to detect significant departures of $\widehat{b}$ from its expected value are given in figure 1 for several values of $r_G, \sigma_{u_1}, \sigma_{u_2}$ and $a_1 = a_2$ or $a_1 = 0.95$. When $a_1 = a_2$ (*eg*, when sires are being proven simultaneously in 2 environments), numbers of progeny required in each environment to reject the null hypothesis that $r_G = 1$ are minimized at $a_j = 0.7$ to 0.8 for r_G between 0.5 and 0.8. For $a_1 = 0.95$ (*eg*, when proven sires are chosen from environment 1), progeny numbers in environment 2 are minimized with one progeny per sire, but increase little until a_2 exceeds 0.5 to 0.6. Thus relatively efficient designs at $a_1 = 0.95$ would include 35 to 45 sires with 400–500 progeny at $r_G = 0.6$, but 250–400 sires with 1 200–1 300 progeny at $r_G = 0.8$.

Critical numbers required for correlation analysis were similar to those for regression analysis at low accuracies and r_G, but lower at higher accuracies due to asymptotic declines in the width of confidence intervals as expected r increased. The ratio of the critical number of sires for correlation analysis to critical number for regression analysis (SRAT) was predictable ($R^2 = 0.983$) as a function of $q = a_1 a_2$ and r_G such that SRAT $= 1.115 - 0.101\,q - 0.667\,q^2 - 0.161\,r_G$. This ratio adjustment can be applied directly to values in figure 1 to approximate critical sire and progeny numbers for correlation analysis.

The above derivations assume that accuracies are calculated correctly in both environments. Under BLUP, accuracies of $\widehat{u}_i$ for non-inbred animals are given by $(1 - C_{ii}/\sigma_u^2)^{.5}$ where C_{ii} is the ith diagonal element of $\mathbf{C}_{22}$, the prediction error covariance matrix of $\widehat{u}$ (Henderson, 1973). If the model is complete and properly parameterized, accuracies are expected to equal correlations between actual and predicted BV. In most applications, $\widehat{u}$ is derived by iterative solution of Henderson's (1963) mixed model equations (MME) rather than by direct inversion. Diagonal
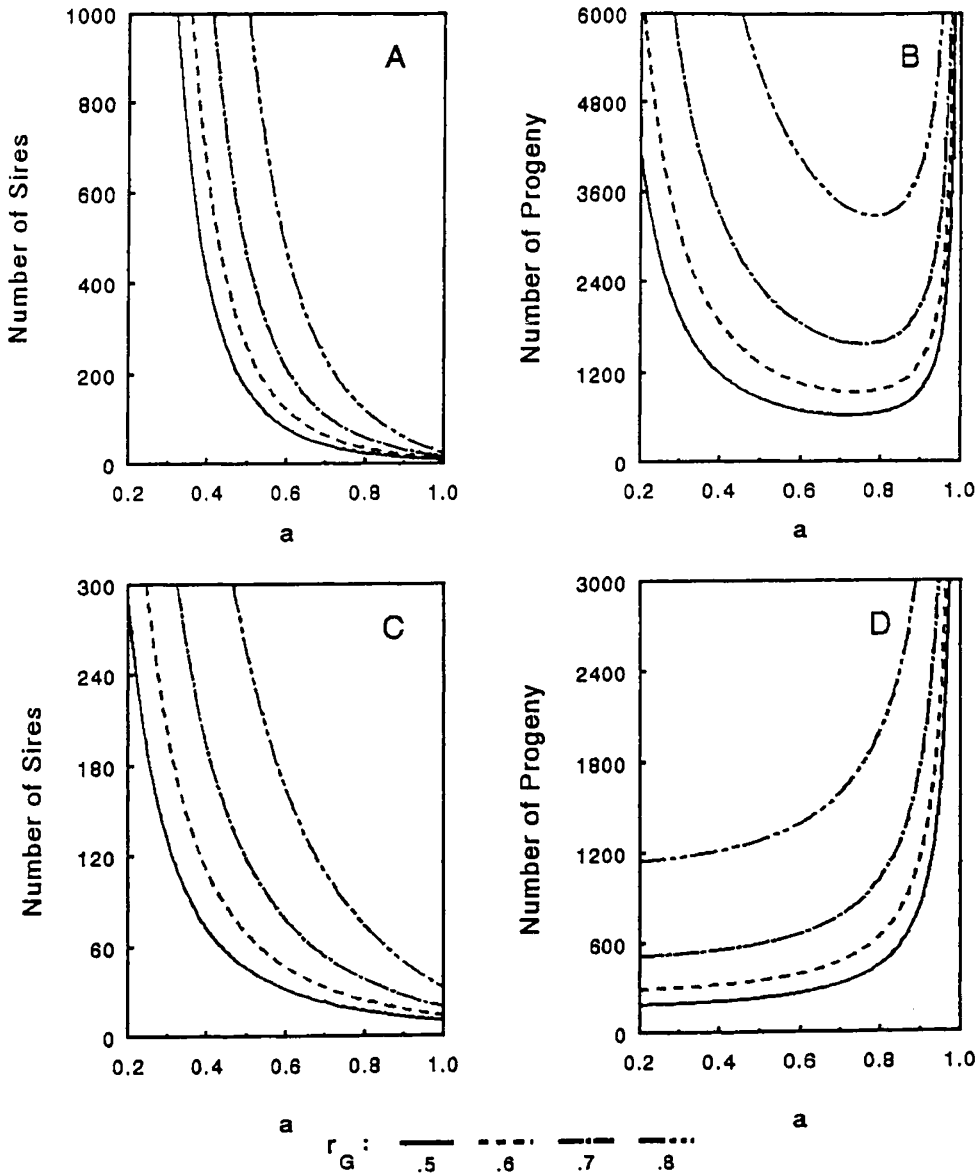
Fig 1. Numbers of sires (N_S) and total progeny in environment 2 (N_T) required for a one-tailed test to detect a significant departure at the 5% level of the observed regression of predicted breeding value (BV) in environment 2 on predicted BV in environment 1 from its expected value under the null hypothesis that the genetic correlation (r_G) between performance in the 2 environments is 1. Results are presented for a range of values for true r_G and as a function of accuracy in environment 2 (a). N_T is calculated as $N_S \cdot NPS$ where NPS = number of progeny/sire in environment 2 is $a^2\lambda/(1-a^2)$. $\lambda = 15$ is the ratio of residual to sire variance. Accuracy in environment 1 is assumed to be equal to a (plates A and B) or to 0.95 (plates C and D).

elements of $\mathbf{C}_{22}$ are approximated but off-diagonal elements of $\mathbf{C}_{22}$ are usually not estimated. To date, no completely satisfactory procedures to obtain diagonal elements of $\mathbf{C}_{22}$ exist. Alternative methods have been presented by Van Raden and Freeman (1985), Greenhalgh *et al* (1986), Robinson and Jones (1987), Meyer (1989) and Van Raden and Wiggans (1991). Evaluation of procedures to estimate accuracy is beyond the scope of this study, but the assumption that accuracies are estimated correctly is critical to the discussion.

Effects of departures from the ideal conditions described above will now be discussed.

Effects of non-random selection from environment 1

Let sires be non-randomly selected based on $\widehat{u}_1$ and accuracies be constant within each environment. Let unselected population variances, covariances, correlations and regressions be symbolized by σ^2, σ, r and b, respectively, and let V, Cov, $Corr$ and $Regr$ respectively represent observed values for some sample from the population. For truncation selection on $\widehat{u}_1$,

$$V(\widehat{u}_1) = (1 - w)\sigma^2_{u_1}$$

where $w = 1 - V(\widehat{u}_1)/\sigma^2_{u_1}$ (Robertson, 1966). For directional truncation selection, $w = i(i - x)$ and for divergent truncation selection $w = -ix$ where i is the standardized selection differential (Becker, 1984) and x is the truncation point on a standard normal curve (Snedecor and Cochran, 1967) corresponding to random selection of sires from the upper or lower fraction, p, of the $\widehat{u}_1$ distribution for directional selection or from the upper and lower fraction $p/2$ for divergent selection. Also:

$$V(\widehat{u}_2) = (1 - a_1^2 r_G^2 a_2^2 w)a_2^2 \sigma_{u_2}^2 \qquad [1]$$

$$Cov(\widehat{u}_1 \widehat{u}_2) = a_1^2 r_G a_2^2 (1 - w)\sigma_{u_1}\sigma_{u_2} \qquad [2]$$

$$Corr(\widehat{u}_1 \widehat{u}_2) = a_1 r_G a_2 \sqrt{\frac{1 - w}{1 - a_1^2 r_G a_2^2 w}} \qquad [3]$$

$$Regr(\widehat{u}_2 \widehat{u}_1) = a_2^2 r_G (\sigma_{u_2}/\sigma_{u_1}) \qquad [4]$$

(Hill, 1970; Johnson and Kotz, 1970; Robertson, 1977). The observed correlation is thus biased by selection but the observed regression is not, and the deviation of $Regr\,(\widehat{u}_2 \widehat{u}_1)$ from its expected value provides a test of the hypothesis that $r_G = 1.0$.

If selection is non-random but not clearly directional or divergent or not based on truncation, additional complications arise. To account for such selection, let $V(\widehat{u}_1)$ be calculated for the selected sample and define w empirically as the observed value $1 - V(\widehat{u}_1)/\sigma^2_{u_1}$. Use of this empirical value of w to predict r_G using equation [3] was evaluated by computer simulation. Predicted BV for the ith sire in environment 1 was simulated as:

$$\widehat{u}_{1i} = \Theta_{1i} a_1 \sigma_{u_1}$$

where Θ_{1i} is a random normal deviate (SAS, 1985), $\sigma_{u_1}^2 = 315$ and $a_1 = 0.7$. Predicted BV in environment 2 were then simulated for $a_2 = 0.7$ and $\sigma_{u_1}^2 = \sigma_{u_2}^2$ as:

$$\widehat{u}_{2i} = r_G a_2^2 \widehat{u}_{1i} + \Theta_{2i}\sigma_{\widehat{u}_2|\widehat{u}_1}$$

where:

$$\sigma_{\widehat{u}_2|\widehat{u}_1}^2 = (1 - a_1^2 r_G^2 a_2^2)a_2^2\sigma_{u_2}^2.$$

Three selection scenarios (SS) were considered:

SS1. 80% divergent, 20% random: 80% of the bulls chosen such that $|x| > 1.282(i = 1.755)$ and 20% chosen at random;

SS2. 50% high, 50% random: 50% of the bulls had $x > 0.842$ ($i = 1.400$);

SS3. 50% high, 50% stabilizing: $|x| < 0.5$ for 50% of the bulls and $x > 1.282$ for 50% of the bulls.

Each scenario was repeated for $r_G = 1.0$ or 0.5 and replicated 10 times. Each replicate contained 5 000 selected animals.

Agreement between predicted and simulated values of $V(\widehat{u}_2)$ and $Corr(\widehat{u}_1\widehat{u}_2)$ (table I) was within theoretical 95% confidence limits of the expected value (Snedecor and Cochran, 1967). Thus equation [3] predicted $Corr(\widehat{u}_1\widehat{u}_2)$ satisfactorily in bulls selected non-randomly on $\widehat{u}_1$ with fixed accuracy a_1.

Table I. Mean simulated values of $V(\widehat{u}_1)$ and w and simulated and expected values (in parentheses) of $V(\widehat{u}_2)$ and $Corr(\widehat{u}_1\widehat{u}_2)$ for each selection scenario and for $r_G = 1.0$ and 0.5[a].

Selection scenario[b]	$r_G = 1.0$				$r_G = 0.5$			
	$V(\widehat{u}_1)$	w	$V(\widehat{u}_2)$	$Corr(\widehat{u}_1\widehat{u}_2)$	$V(\widehat{u}_1)$	w	$V(\widehat{u}_2)$	$Corr(\widehat{u}_1\widehat{u}_2)$
80D, 20R	430.7	−1.790	219.0 (220.7)	0.686 (0.684)	432.2	−1.800	170.6 (171.0)	0.390 (0.389)
50H, 50R	169.9	−0.101	155.9 (158.1)	0.507 (0.508)	167.5	−0.085	155.0 (155.1)	0.252 (0.255)
50H, 50S	137.9	0.106	151.6 (150.4)	0.466 (0.469)	138.5	0.103	154.4 (153.4)	0.231 (0.233)

[a] Each entry is the mean of 10 replications of 5 000 observations each; [b] see text for definition of selection scenarios.

With selected sires, the expected $V(\widehat{b})$ is:

$$V(\widehat{b}) = [1 - Corr^2(\widehat{u}_1\widehat{u}_2)]V(\widehat{u}_2)/[(m-1)V(\widehat{u}_1)]$$

using values from equations [1] and [3]. The SD of $\widehat{b}$ is inversely proportional to $\sqrt{1-w}$ and varies from 48 to 243% of its value when $w = 0$ as w varies from -3.39 (divergent selection from the top and bottom 5% of the population) to 0.83 (selection from the top 10% of the population). Sample sizes to detect significant

departures of $Regr(\widehat{u}_2\widehat{u}_1)$ from its expectation using selected sires can be derived from figure 1 by dividing sire and total animal numbers by $1 - w$.

Effects of variation among animals in accuracy of predicted BV

Calo *et al* (1973) and Blanchard *et al* (1983) derived the expected correlation between predicted BV for 2 traits when BV for each trait were estimated in separate single-trait analyses and individuals differed in accuracy of BV prediction as $Corr(\widehat{u}_1\widehat{u}_2) = r_G a_{12}$ for:

$$a_{12} = \sum_{i=1}^{m} a_{1i}^2 a_{2i}^2 \Big/ \sqrt{\left(\sum_{i=1}^{m} a_{1i}^2\right)\left(\sum_{i=1}^{m} a_{2i}^2\right)} \qquad [5]$$

(see *Appendix*) and recommended using this expression to estimate r_G from observed $Corr(\widehat{u}_1\widehat{u}_2)$. Similarly:

$$Regr(\widehat{u}_2\widehat{u}_1) = \left(r_G \sum_{i=1}^{m} a_{1i}^2 a_{2i}^2\right) \Big/ \left(\sum_{i=1}^{m} a_{1i}^2\right) \qquad [6]$$

Taylor (1983) criticized equation [5] as unstable, however, asserting that it may yield estimates of r_G that are outside the parameter space, and presented assumptions required to allow estimation of r_G with this equation. Taylor (1983) concluded that, if all assumptions are met, equation [5] is appropriate to estimate r_G so long as the a_j^2 are derived from MME as $(1 - C_{ii}/\sigma_u^2)$.

The equations of Calo *et al* (1973) and Blanchard *et al* (1983) do not consider selection on $\widehat{u}_1$. With selection, equation [5] appears appropriate to estimate $Corr(\widehat{u}_1\widehat{u}_2)$ for the selected sample, but not within the unselected population. Equations [3] and [5] could, however, perhaps be combined to give the expected correlation in a selected sample of animals with variable accuracy as:

$$Corr(\widehat{u}_1\widehat{u}_2) = r_G a_{12} \sqrt{\frac{1 - w}{1 - r_G^2 a_{12}^2 w}} \qquad [7]$$

To evaluate equation [7], several accuracy scenarios (AS) were considered by simulating samples from the $\widehat{u}_1$ distribution.

AS1: 20 000 animals from the upper 10% of the $\widehat{u}_1$ distribution. a_1 varied uniformly over the interval 0.7 to 0.95.

AS2: 20 000 animals from the upper and lower 10% of the $\widehat{u}_1$ distribution. a_1 varied uniformly over the interval 0.7 to 0.95.

AS3: 10 000 animals from the upper 10% of the $\widehat{u}_1$ distribution with a_1 uniformly distributed over the interval 0.7 to 0.95 and 10 000 animals selected from the lower 10% of the distribution with a_1 uniformly distributed over the interval 0.5 to 0.7.

AS4: 10 000 animals from the upper 10% of the $\widehat{u}_1$ distribution with a_1 uniformly distributed over the interval 0.7 to 0.95 and 10 000 animals selected from

the bottom 80% of the distribution with a_1 uniformly distributed over the interval 0.5 to 0.7.

AS5: 15 000 animals from the upper 10% and 5 000 animals from the lower 80% of the $\widehat{u}_1$ distribution with a_1 uniformly distributed over the interval 0.7 to 0.95.

AS6: 5 000 animals from the upper 10% and 15 000 animals from the bottom 80% of the $\widehat{u}_1$ distribution with a_1 uniformly distributed over the interval 0.5 to 0.995.

AS7: 10 000 animals from the upper 10% of the $\widehat{u}_1$ distribution with a_1 uniformly distributed over the interval 0.795 to 0.995 and 10 000 animals from the lower 80% of the distribution with a_1 uniformly distributed over the interval 0 to 0.50.

a_2 was uniformly distributed over the interval 0.5 to 0.7 for all scenarios. Each scenario was repeated for $r_G = 0.5$ or 1.0 and replicated (table II). Empirical calculation of w requires use of $\sigma_{u_1}^2$, which varies with accuracy. Simulated values of $\widehat{u}_1$ were thus standardized by dividing by $a_{1i}\sigma_{u_1}$ and empirical w calculated as $1 - V(\widehat{u}_1)$ using standardized $\widehat{u}_1$.

Table II. Observed and predicted mean correlation and regression coefficient for each accuracy scenario with genetic correlations (r_G) of 1 or 0.5[a].

Accuracy scenario[b]	r_G	Mean correlation		Mean regression	
		Predicted	Observed	Predicted	Observed
1	1.0	0.230	0.240*	0.363	0.361
	0.5	0.105	0.116*	0.182	0.187
2	1.0	0.720	0.715*	0.363	0.363
	0.5	0.422	0.420	0.182	0.181
3	1.0	0.659	0.650*	0.364	0.363
	0.5	0.374	0.370*	0.182	0.181
4	1.0	0.509	0.524*	0.363	0.362
	0.5	0.263	0.275*	0.182	0.182
5	1.0	0.517	0.514	0.363	0.363
	0.5	0.262	0.261	0.182	0.181
6	1.0	0.507	0.503*	0.363	0.360
	0.5	0.260	0.259	0.182	0.181
7	1.0	0.471	0.503*	0.363	0.365
	0.5	0.258	0.264*	0.182	0.182

[a] Each entry is the mean of 10 replicates of 20 000 observations each; [b] see text for definition of accuracy scenarios; * observed and predicted values differ ($P < 0.05$).

For all accuracy scenarios, observed $Regr(\widehat{u}_2\widehat{u}_1)$ agreed closely with predicted values from equation [6] (table II). Observed values of $Corr(\widehat{u}_1\widehat{u}_2)$ were usually also close to expectations from equation [7], but with some systematic departures from expectations. For directional selection (AS1), the mean observed $Corr(\widehat{u}_1\widehat{u}_2)$ was slightly but significantly larger than predicted (by 0.010 ± 0.002 for both r_G).

Thus equation [7] produced a small negative bias under directional selection with variable accuracy. This result was confirmed by producing 10 more replicates at $r_G = 1$; the results were identical.

For divergent selection, differences between observed and predicted correlations were again small, but sometimes significant and now negative for AS2, 3, 5 and 6, ranging from -0.001 to -0.008 (± 0.002). However, with both non-symmetrical selection from high and low groups and different accuracy distributions between groups (AS4 and AS7), observed correlations were considerably larger than predicted, especially for $r_G = 1$ (table II). The appendix shows exact expectations for correlations and regressions involving $\hat{u}_1$ and $\hat{u}_2$ under non-random selection from environment 1 and variable accuracies within environments. Correlation between means and accuracies of divergently selected groups violate some of the assumptions used to derive equation [7] and presumably account for the departures from predicted values in AS4 and AS7.

Equation [7] thus produced slightly biased predictors of $Corr(\hat{u}_1\hat{u}_2)$ but still appears useful, especially when exact selection rules are unknown. However, biases in predicted values of $Corr(\hat{u}_1\hat{u}_2)$ in equation [7] will be multiplied by the inverse of the coefficient of r_G in equation [7] to estimate r_G. Potential bias in r_G thus is larger with lower a_j or more directional selection. If $V(\hat{u}_i)$ is larger than expected from random selection and greater precision than that provided by equation [7] is desired, *Appendix* equations can be used.

The expected correlation between $\hat{u}_1$ and $\hat{u}_2$ thus depends on the distribution of accuracies within each environment, the selection applied on $\hat{u}_1$ (quantified by w) and r_G. To evaluate net effects of these variables, values of $Corr(\hat{u}_1\hat{u}_2)$ from equation [7] were calculated for $r_G = 1$ and when a_{12} varied from 0.10 to 0.90 and w varied from -3.3 to 0.9 (table III). For $r_G = 1.0$ and $a_{12} = 0.90$, $Corr(\hat{u}_1\hat{u}_2)$ varied from 0.55 to 0.97 due to selection, although r_G exceeded 0.79 so long as directional selection was not intense (w 6 0.6). For $w = 0$ and $r_G = 1.0$, $Corr(\hat{u}_1\hat{u}_2)$ equalled a_{12}, but still varied from 0.10 to 0.90, depending on observed values of a_1 and a_2.

Effects of relationships

Previous results assume independence of predicted BV within each environment. However, relationships among animals lead to covariances among predicted BV within and across environments. If $\hat{u}_1$ and $\hat{u}_2$ are predicted by BLUP, covariances among predicted BV within environments also arise from estimation of fixed effects. These covariances affect expectations of both $Corr(\hat{u}_1\hat{u}_2)$ and $Regr(\hat{u}_2\hat{u}_1)$. Their impact is difficult to generalize, depending upon the extent and nature of relationships in the data and the distribution of records among fixed effect classes.

Covariances among predicted BV associated with relationships and estimation of fixed effects arise simultaneous in BLUP solutions to MME, but effects of relationships alone can be seen under selection index, or best linear prediction (BLP), assumptions of known mean and variance for both u_1 and u_2. In that case:

$$\hat{\mathbf{u}}_j = \mathbf{H}_j' \mathbf{V}_j^{-1} \mathbf{y}_j$$

where $\hat{\mathbf{u}}_j$ is a vector of breeding value predictions for environment j and $\mathbf{y}_j$ is the data vector in environment j with covariance matrix $\mathbf{V}_j$. $\mathbf{H}_j$ is the covariance

Table III. Expected correlation between predicted breeding values in 2 environments as a function of accuracies (a_{12}) and extent of selection in environment 1 (w)[a].

	a_{12}				
w	0.10	0.30	0.50	0.70	0.90
− 3.3	0.204	0.546	0.767	0.897	0.974
− 3.0	0.197	0.532	0.756	0.891	0.972
− 2.7	0.190	0.518	0.743	0.883	0.970
− 2.4	0.182	0.502	0.729	0.875	0.967
− 2.1	0.174	0.484	0.713	0.865	0.964
− 1.8	0.166	0.466	0.695	0.854	0.961
− 1.5	0.157	0.445	0.674	0.840	0.956
− 1.2	0.147	0.423	0.650	0.824	0.951
− 0.9	0.137	0.398	0.623	0.804	0.943
− 0.6	0.126	0.370	0.590	0.778	0.934
− 0.3	0.114	0.338	0.550	0.745	0.920
0	0.100	0.300	0.500	0.700	0.900
0.3	0.084	0.254	0.435	0.634	0.865
0.6	0.063	0.195	0.343	0.527	0.794
0.9	0.032	0.099	0.180	0.296	0.547

[a] Tabular entries are derived from text equation [7] for a genetic correlation of 1.0. $w = 1$ minus the ratio of observed to expected variance in predicted BV in environment 1, and a_{12} is from text equation [5].

matrix between $\mathbf{y}_j$ and $\mathbf{u}'_j$. For non-inbred animals, $\mathbf{H}_j = \mathbf{Z}_j\mathbf{G} = \mathbf{Z}_j\mathbf{A}\sigma^2_{u_j}$ where $\mathbf{Z}_j$ is the incidence matrix relating $\mathbf{y}_j$ to $\mathbf{u}_j$, and $\mathbf{G}$ and $\mathbf{A}$ are additive covariance and numerator relationship matrices, respectively, for animals in $\mathbf{u}_j$. The covariance matrix of $\widehat{\mathbf{u}}_j(\mathbf{Q}_j)$ and the covariance matrix between $\widehat{\mathbf{u}}_1$ and $\widehat{\mathbf{u}}'_2$ ($\mathbf{Q}_{12}$) are thus:

$$\mathbf{Q}_j = \mathbf{H}'_j\mathbf{V}_j^{-1}\mathbf{H}_j$$
$$\mathbf{Q}_{12} = \mathbf{H}'_1\mathbf{V}_1^{-1}Cov(\mathbf{y}_1\mathbf{y}'_2)\mathbf{V}_2^{-1}\mathbf{H}_2.$$

Expectations of sample variances of $\widehat{u}_j$ ($s^2_{\widehat{u}_j}$) and covariance between $\widehat{u}_1$ and $\widehat{u}_2$ ($s_{\widehat{u}_1\widehat{u}_2}$) are functions of elements of $\mathbf{Q}_j$ and $\mathbf{Q}_{12}$:

$$s^2_{\widehat{u}_j} = [mtr(\mathbf{Q}_j) - sum(\mathbf{Q}_j)]/[m(m-1)] \qquad [8]$$

$$s_{\widehat{u}_1\widehat{u}_2} = [mtr(\mathbf{Q}_{12}) - sum(\mathbf{Q}_{12})]/[m(m-1)] \qquad [9]$$

where tr is the trace of the matrix and sum is them sum of all elements. If animals recorded in the 2 environments resemble one another only because of relationships to animals sampled from environment 1,

$$Cov(\mathbf{y}_1\mathbf{y}'_2) = \mathbf{Z}_1\mathbf{A}\mathbf{Z}'_2\sigma_{u_1}\sigma_{u_2}r_G \Rightarrow \qquad [10]$$

$$\mathbf{Q}_{12} = \mathbf{Q}_1\mathbf{A}^{-1}\mathbf{Q}_2(r_G/\sigma_{u_1}\sigma_{u_2}) \qquad [11]$$

This assumption is warranted if animals are evaluated using unrelated populations of mates in the 2 environments but may not be correct if mates are potentially related across environments. If selected animals are likewise unrelated, and relationships among recorded animals within each environment arise only through relationships to selected animals, $\mathbf{Q}_j$ will be diagonal with elements $a_{ij}^2 \sigma_{u_j}^2$ for the ith animal in the jth environment, $\mathbf{A}$ is an identity matrix of size m and $\mathbf{Q}_{12}$ is diagonal with elements $a_{1i}^2 a_{2i}^2 r_G \sigma_{u_1} \sigma_{u_2}$. Sample correlation and regression coefficients then have expectations equal to those previously discussed.

In most applications, $\mathbf{Q}_j$ are not diagonal and off-diagonal elements are not calculated due to the nature and size of $\mathbf{V}_j^{-1}$. Thus explicit consideration of off-diagonal elements of $\mathbf{Q}_j$ may not be possible unless the data set is small or highly structured. However, if the accuracy of all $\widehat{u}_{ij}$ approaches 1, $\mathbf{Q}_j$ approaches $\mathbf{A}\sigma_{u_j}^2$ (ie, $\sigma_{u_j}^2 \to \sigma_{u_j}^2$) and $\mathbf{Q}_{12}$ approaches $\mathbf{A}_{r_G}\sigma_{u_1}\sigma_{u_2}$ such that $Corr(\widehat{u}_1\widehat{u}_2) = r_G$ and $Regr(\widehat{u}_2\widehat{u}_1) = r_G \sigma_{u_2}/\sigma_{u_1}$.

A more realistic situation is one in which sires from environment 1 have a_1 approaching 1 but are evaluated in environment 2 with $a_2 < 1$. In this case, $\mathbf{Q}_1 = \mathbf{A}\sigma_{u_1}^2$, $\mathbf{Q}_2 = \mathbf{H}_2'\mathbf{V}_2^{-1}\mathbf{H}_2$, and $\mathbf{Q}_{12} = \mathbf{Q}_2(r_G \sigma_{u_1}/\sigma_{u_2})$. These quantities, and associated expected sample variances and covariances, could be obtained if the size or structure of the data allows calculation of all elements of $\mathbf{Q}_2$ and would allow derivation of an exact predicted value for $Corr(\widehat{u}_1\widehat{u}_2)$.

A small example will demonstrate the impact of relationships on $Corr(\widehat{u}_1\widehat{u}_2)$ and $Regr(\widehat{u}_2\widehat{u}_1)$, which are equal in these examples. Let $h^2 = 0.25$, $\sigma_{u_1} = \sigma_{u_2} = 1$ and $r_G = 1$. Let 3 sires produce 8 progeny each in each of 2 environments. If sires are unrelated and progeny are related only through the sires, sample variances and covariances involving $\widehat{u}_1$ and $\widehat{u}_2$ equal expected population values, and $Corr(\widehat{u}_1\widehat{u}_2) = a_1 a_2$. If all 3 sires are full sibs, $\mathbf{Q}_j$ and $\mathbf{Q}_{12}$ are no longer diagonal and $a_1 = a_2 = 0.643$. The expected $Corr(\widehat{u}_1\widehat{u}_2)$ is 0.211 versus $a_1 a_2 = 0.414$. If sires are half sibs, $Corr(\widehat{u}_1\widehat{u}_2) = 0.286$ versus $a_1 a_2 = 0.365$, reducing bias by $\approx$ one half as relationships among sires decline. Still, with many close relationships among sires, $a_1 a_2$ may considerably overestimate the expected correlation.

If only 2 of the 3 sires are full sibs, bias is reduced. $a_1 = a_2 = 0.621$ for related sires and 0.590 for the unrelated sire, and $a_{12} = 0.374$ (equation [5]). Now a_{12} overestimates the observed correlation by only 3.9% (0.374 versus 0.360). Thus if sampled animals represent a reasonable number of unrelated families, $\mathbf{Q}_j$ and $\mathbf{Q}_{12}$ are correspondingly sparse and little bias in $Corr(\widehat{u}_1\widehat{u}_2)$ is expected.

Turning to effects of relationships within environments, let the 3 sires be unrelated but cross-classified within each environment with only 2 dams. In that case, $a_1 = a_2 = 0.566$ and $Corr(\widehat{u}_1 u_2) = 0.364$ which is 13% larger than $a_1 a_2 = 0.321$. However, under the more realistic assumption of cross-classification with 2 maternal grandsires, $Corr(\widehat{u}_1\widehat{u}_2) = 0.352$ versus $a_1 a_2 = 0.338$, yielding little bias. When sires were cross-classified with 8 dams, $Corr(\widehat{u}_1\widehat{u}_2) = 0.364$ versus $a_1 a_2 = 0.349$, again yielding little bias. These examples suggest selection of widely proven, lowly related animals from environment 1 followed by evaluation in environment 2 using a broad sample of mates.

Effects of covariances arising from estimation of fixed effects

If fixed effects are estimated simultaneously with BLUP of $\widehat{u}_j$, Mallinckrodt (1990) noted that off-diagonal elements of $\mathbf{Q}_j$ and $\mathbf{Q}_{12}$ are not zero, even in the absence of relationships. By BLUP:

$$
\begin{aligned}
\widehat{u_j} &= \mathbf{H}'_j \mathbf{V}_j^{-1}(\mathbf{y}_j - \mathbf{X}_j \widehat{\boldsymbol{\beta}}_j) \\
&= \mathbf{H}'_j \mathbf{V}_j^{-1}[\mathbf{I} - \mathbf{X}_j(\mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j)^{-1} \mathbf{X}'_j \mathbf{V}_j^{-1}]\mathbf{y}_j \\
&= \mathbf{H}'_j \mathbf{V}_j^{-1}[\mathbf{I} - \mathbf{P}_j \mathbf{V}_j^{-1}]\mathbf{y}_i
\end{aligned}
$$

for $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j)^{-1} \mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{y}_j$ and $\mathbf{P}_j = \mathbf{X}_j(\mathbf{X}'_j \mathbf{V}_j^{-1} \mathbf{X}_j)^{-1} \mathbf{X}'_j$ and where $\boldsymbol{\beta}$ is the vector of fixed effects with incidence matrix $\mathbf{X}$.

The covariance matrix of $\widehat{u}_j$ is:

$$
\begin{aligned}
\mathbf{Q}_j &= \mathbf{H}'_j \mathbf{V}_j^{-1}[\mathbf{I} - \mathbf{P}_j \mathbf{V}_j^{-1}]\mathbf{V}_j[\mathbf{I} - \mathbf{P}_j \mathbf{V}_j^{-1}]'\mathbf{V}_j^{-1}\mathbf{H}_j \\
& \mathbf{H}'_j \mathbf{V}_j^{-1}[\mathbf{I} - \mathbf{P}_j \mathbf{V}_j^{-1}]\mathbf{H}_j
\end{aligned}
$$

composed of a term due to BLP relationships minus a term due to estimation of fixed effects. If relationships across environments arise only through sires sampled from environment 1 such that $Cov(\mathbf{y}_1 \mathbf{y}'_2)$ is given by equation [10], $\mathbf{Q}_{12}$ is still given by equation [11], but regardless of the relationship structure of the data, $\mathbf{Q}_j$ now approaches a diagonal matrix only as a_j approach 1. Also, for a given number of progeny, a_{ij} will be less for BLUP than for BLP and depends on the number of sires and their distribution among fixed effect classes. The above solutions are identical to those obtained from MME (Henderson, 1963, 1984) such that:

$$
\mathbf{Q}_j = \mathbf{G}_j - \mathbf{C}_{22_j}.
$$

The impact of fixed effect estimation on $Corr(\widehat{u}_1 \widehat{u}_2)$ can be seen most readily using MME for a sire model without relationships among animals in $\mathbf{u}$ and where $\boldsymbol{\beta}$ includes only contemporary group effects. Note that in all remaining examples, $Corr(\widehat{u}_1 \widehat{u}_2) = Regr(\widehat{u}_2 \widehat{u}_1)$ when $a_1 = a_2$. For such a model, after absorption of fixed effects into $\widehat{\mathbf{u}}$ equations and factoring of residual variance from both sides of the equation, the coefficient matrix for $\widehat{\mathbf{u}}$ has:

$$
n_{i.} - \sum_{k=1}^{q}(n_{ik}^2/n_{.k}) + \lambda \text{ on the diagonals, and}
$$

$$
- \sum_{k=1}^{q}(n_{ik}n_{i'k}/n_{.k}) \text{ on the off-diagonals}
$$

(Do, 1991) where λ is the ratio of residual to sire variance and $n_{i.}, n_{.k}$ and n_{ik} are numbers of records for sire i, contemporary group k (of g) and sire $\times$ contemporary group subclass ik, respectively. For balanced data, $n_{ik} = n$ for all i and k, diagonals reduce to $[gn(m-1)/m + \lambda][vs(gn + \lambda)$ for BLP] and

off-diagonals reduce to $(-gn/m)$ [vs 0 for BLP]. The corresponding inverse of the coefficient matrix has diagonal elements of $(gn + m\lambda)/[m(gn + \lambda)]$ and off-diagonals of $gn/[m\lambda(gn + \lambda)]$ (Searle, 1966). $\mathbf{Q}$ will have corresponding diagonal and off-diagonal elements of $gn(m - 1)/[m(gn + \lambda)] = a^2$ and $-gn/[m(gn + \lambda)]$, respectively. s^2 is thus $gn/(gn + \lambda) = a^2[(m - 1)/m]$. If design matrices are the same for both environments and $r_G = 1$, $\mathbf{Q}_{12}$ will have diagonal elements of $g^2n^2(m-1)/[m(gn+\lambda)^2] = a^4m/(m-1)$ and off-diagonals of $-g^2n^2/[m(gn+\lambda)^2] = -a^4m/(m - 1)$ to give $s_{12} = g^2n^2/(gn + \lambda)^2 = a^4m^2/(m - 1)^2$ and $Corr(\widehat{u}_1\widehat{u}_2) = gn/(gn + \lambda) = a^2m/(m - 1)$. With balanced designs, $Corr(\widehat{u}_1\widehat{u}_2)$ has the same expectation under both BLUP and BLP, but accuracies are lower under BLUP such that a_1a_2 from BLUP underestimates expected $Corr(\widehat{u}_1\widehat{u}_2)$. The extent of bias is proportional to $m/(m-1)$ and decreases from 20% at $m = 5$ to 11% at $m = 10$ and 2.6% at $m = 40$. This expectation is maintained if design matrices differ between environments provided designs are balanced within each environment.

The situation is more complicated for unbalanced designs, but general conclusions are similar in that the number of sires compared as contemporaries needs to be large enough to minimize confounding between sire BV predictions and fixed effects estimates. Otherwise, BLUP accuracies are reduced and their product underestimates expected $Corr(\widehat{u}_1\widehat{u}_2)$. For example, consider a block of 8 sires with progeny distributed over 4 contemporary groups (eg, 4 yr of an experimental evaluation in some environment) as shown in table IV. The size of the experiment may be varied by increasing the number of sire blocks (to 16, 24, etc, sires), by varying the number of progeny per sire and contemporary group (n), by replicating the sire block over additional contemporary groups (8, 12, 16, etc), or by a combination of these approaches. The same design is assumed for each environment.

Table IV. Design of a sire block to evaluate effects of unbalanced data on expected correlation between predicted breeding values in 2 environments.

Sire	Contemporary group			
	1	*2*	*3*	*4*
A	X			X
B	X			X
C	X	X		
D	X	X		
E		X	X	
F		X	X	
G			X	X
H			X	X

Define bias (fig 2) as the difference between the expected $Corr(\widehat{u}_1\widehat{u}_2)$ calculated from equations [8] and [9] and the product a_1a_2 which is constant for all sires in this design. With $n = 6$ and only 8 sires, bias was relatively large with $Corr(\widehat{u}_1\widehat{u}_2) = 0.41$ vs $a_1a_2 = 0.35$. Bias decreased as number of sire blocks increased and was < 0.03 with 24 sires (12/contemporary group). Doubling n or replicating sire blocks across more contemporary groups (not shown) did little to change the

pattern of bias. Expected values of $Corr(\widehat{u}_1\widehat{u}_2)$ were also compared to the product of BLP accuracies calculated ignoring contemporary group effects. The product of BLP accuracies overestimated $Corr(\widehat{u}_1\widehat{u}_2)$ as shown by negative bias in figure 2, but the product of BLP accuracies was superior to the product of BLUP accuracies as an estimator of $Corr(\widehat{u}_1\widehat{u}_2)$.
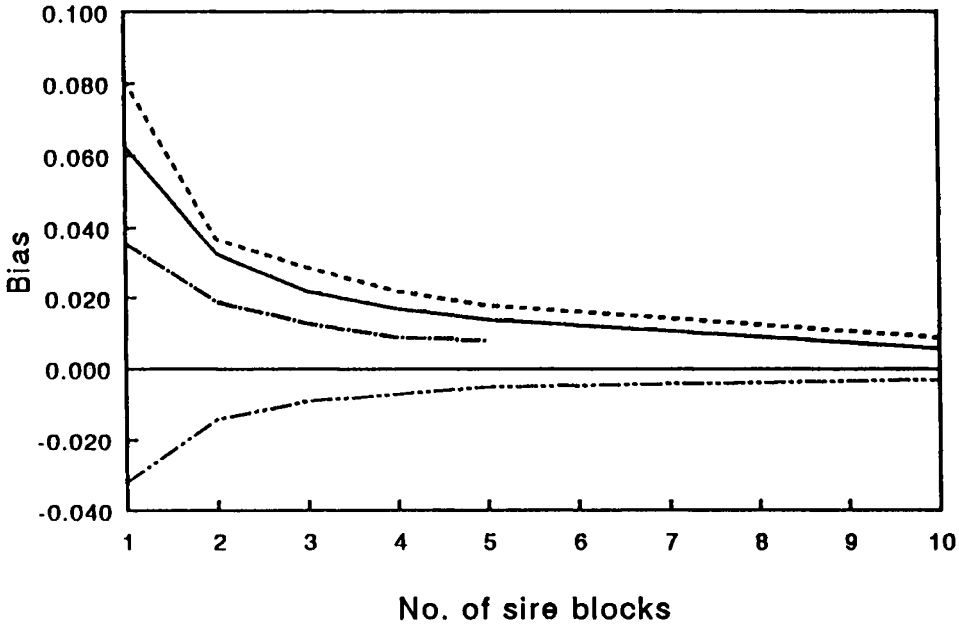

No. of sire blocks

Fig 2. Effect of number of sire blocks (table IV) on bias from estimation of the expected correlation between predicted breeding values in environments 1 and 2 $[Corr(\widehat{u}_1\widehat{u}_2)]$ from the product of BLP or BLUP accuracies (a_1a_2) in the environments. Biases are calculated as $Corr(\widehat{u}_1\widehat{u}_2) - a_1a_2$. Bias associated with BLUP a_j are shown above the X axis for 6 progeny (center line) or 12 progeny (upper line) per sire and contemporary group or for 6 progeny with addition of 2 non-connecting sires in each contemporary group (lower line). Bias associated with BLP a_j are shown below the X axis for 6 progeny per sire and contemporary group.

Design like that in table IV can be used under experimental conditions, but are less feasible when sires are compared on cooperator farms. It particular, use of large numbers of experimental sires on individual farms may not be feasible. Instead, sires from environment 1 may be tested together on several farms in a loosely connected design but with only a few sires represented on any one farm. If only data from introduced sires are used in the evaluation, considerable bias in $Corr(\widehat{u}_1\widehat{u}_2)$ may result. However, this bias can be reduced if introduced sires are evaluated with sires represented only in environment 2 and data from all sires are included in the evaluation. To demonstrate this effect, two additional sires were added to each contemporary group in table IV to give 16 sires/sire block with $n = 6$. Added sires produced progeny in only one contemporary group. Bias in

resulting values of $Corr(\hat{u}_1\hat{u}_2)$ for introduced sires was reduced (fig 2). Thus if sires from one environment are introduced into another and if evaluation occurs primarily in cooperator herds, sires should be evaluated in contemporary groups containing reasonably large numbers of sires (either introduced or native) to increase precision of estimates of contemporary group effects, and data from all sires should be included in the evaluation.

CONCLUSIONS

Interpretation of correlations between predicted BV in different environments is not straightforward. Expectations of such correlations are influenced by accuracy of evaluation of animals in both environments, by selection of animals chosen for evaluation, by relationships among chosen animals and by the design of the evaluation in both environments. If animals are chosen from environment 1 for evaluation in environment 2, $Regr(\hat{u}_2\hat{u}_1)$ may be a more useful statistic than $Corr(\hat{u}_1\hat{u}_2)$ because it is unbiased by selection on $\hat{u}_1$. However, both $Regr(\hat{u}_2\hat{u}_1)$ and $Corr(\hat{u}_1\hat{u}_2)$ are biased by covariances among predicted BV within environments. Also, use of proven sires $(a \rightarrow 1.0)$ from environment 1 simplifies interpretations and reduces the number of sires required to attain a specific level of significance for measures of association.

Equations in this paper allow calculation or approximation of expected values of $Corr(\hat{u}_1\hat{u}_2)$ under various sorts of selection and with variable accuracies in each environment (equation [7]). Evidence for selection can be obtained empirically if necessary by comparing observed $\mathbf{V}(\hat{u}_1)$ to its expectation from Blanchard *et al* (1983) (see *Appendix*). Expected values of both $Corr(\hat{u}_1\hat{u}_2)$ and $Regr(\hat{u}_2\hat{u}_1)$ may involve off-diagonal elements of $\mathbf{Q}_j$ matrices which are often not available for BLUP BV predictions. Effects of off-diagonals may be minimized by ensuring that a number of families are represented in both the experimental animals and their mates and by using several (*eg* 8–16) sires per contemporary group. Use of small numbers of sires per contemporary group can lead to considerable underestimation of the expected value of $Corr(\hat{u}_1\hat{u}_2)$ if off-diagonal elements are not considered. Prediction of expected values of $Corr(\hat{u}_1\hat{u}_2)$ and $Regr(\hat{u}_2\hat{u}_1)$ from observed accuracies may be superior when selection index (BLP) rather than BLUP accuracies are used.

A number of potential difficulties in deriving and interpreting $Corr(\hat{u}_1\hat{u}_2)$ have not been explicitly considered. Accuracies are assumed to be properly calculated, even though approximations are normally used and are probably not completely satisfactory. Additive genetic variances must be known for both environments in order to calculate $\hat{u}_1$ and $\hat{u}_2$ correctly and to interpret $Regr(\hat{u}_2\hat{u}_1)$. If sires introduced into environment 2 for evaluation are a selected sample from environment 1 and BLUP evaluations are used, grouping strategies and(or) adjustment of covariances may be required to derive unbiased $\hat{u}_2$. Values of $\sigma_{u_2}^2$ for animals selected from environment 1 will depend both on selection applied and on r_G. Thus results given for correlations involving BLUP predictions of $\hat{u}_j$ are probably strictly correct only for random selection of sires. See Diaz (1992) for additional discussion of effects of selection and grouping on $Corr(\hat{u}_1\hat{u}_2)$ with BLUP predictions. Despite these problems, however, correlations and regressions involving BV predictions in different

environments will often be relatively easy to obtain and, if properly interpreted, can provide information on r_G.

APPENDIX

This appendix addresses general expectations for variances and covariances of $\widehat{u}_1$ and $\widehat{u}_2$ with non-random selection on $\widehat{u}_1$ and variable accuracies in both environments. Sample m animals from environment 1. If sampling is non-random and selection rules are not specified or if accuracies differ, the resulting distributions of $\widehat{u}_1$ and $\widehat{u}_2$ are mixtures of m unique distributions. The sample sum of squares (SS) of $\widehat{u}_1$ is:

$$SS(\widehat{u}_1) = \sum_{i=1}^{m} \widehat{u}_{1i}^2 - \left(\sum_{i=1}^{m} \widehat{u}_{1i}\right)^2 / m.$$

The expected value (E) and variance of $\widehat{u}_{2i}$ with accuracy a_{2i} are:

$$E(\widehat{u}_{2i}) = a_{2i}^2 r_G (\sigma_{u_2}/\sigma_{u_1}) \widehat{u}_{1i}$$

$$\sigma_{\widehat{u}_{21}|\widehat{u}_{li}}^2 = (1 - a_{li}^2 r_G^2 a_{2i}^2) a_{2i}^2 \sigma_{u_2}^2.$$

The resulting expected $SS(\widehat{u}_2)$ and sum of cross-products (SCP) are:

$$SS(\widehat{u}_2) = \sum_{i=1}^{m} \left[E^2(\widehat{u}_{2i}) + \sigma_{\widehat{u}_{2i}|\widehat{u}_{1i}}^2\right] - \left[\sum_{i=1}^{m} E(\widehat{u}_{2i})\right]^2 / m$$

$$= r_g^2(\sigma_{u_2}^2/\sigma_{u_1}^2)\left[\sum_{i=1}^{m} a_{li}^4 \widehat{u}_{1i}^2 - \left(\sum_{i=1}^{m} a_{2i}^2 \widehat{u}_{li}\right)^2 / m\right] +$$

$$\sum_{i=1}^{m} (1 - a_{1i}^2 r_G a_{2i}^2) a_{2i}^2 \sigma_{u_2}^2$$

$$SCP(\widehat{u}_1 \widehat{u}_2) = \sum_{i=1}^{m} E(\widehat{u}_{i1}\widehat{u}_{i2}) - \left[\sum_{i=1}^{m} E(\widehat{u}_{1i})\right]\left[\sum_{i=1}^{m} E(\widehat{u}_{2i})\right] / m$$

$$= r_G(\sigma_{u_2}/\sigma_{u_1})\left[\sum_{i=1}^{m} a_{2i}^2 \widehat{u}_{1i}^2 - \left(\sum_{i=1}^{m} \widehat{u}_{1i}\right)\left(\sum_{i=1}^{m} a_{2i}^2 \widehat{u}_{1i}\right)/m\right].$$

Expected correlation and regression coefficients can be derived from these SS and SCP. If animals are chosen at random, $\widehat{u}_1$ and $\widehat{u}_2$ have mean 0 and expected $Corr(\widehat{u}_1\widehat{u}_2)$ is given by text equation [5] (Blanchard *et al*, 1983).

If selection is non-random but accuracies are constant within environments, formulae for $SS(\widehat{u}_2)$ and $SCP(\widehat{u}_1\widehat{u}_2)$ become consistent with text formulae 2 and 3. For directional selection with constant accuracies, w can be replaced by $i(i - x)$ without loss of generality. For divergent selection, let fraction w_H of the selected animals come at random from the upper fraction ϕ_H of the distribution with

standardized mean and truncation point of i_H and x_H, respectively. Let fraction w_L come from the lower fraction ϕ_L. Let accuracies within high and low groups, and within environment 2, be constant at a_{1H}, a_{1L} and a_2, respectively. $V(\widehat{u}_1)$ within selected groups are $\left[1 - i_H(i_H - x_H)\right]a_{1H}^2\sigma_1^2$ and $\left[1 - i_L(i_L - x_L)\right]a_{1L}^2\sigma_1^2$. The expected value and variance of $\widehat{u}_1$ and $\widehat{u}_2$, and their expected covariance are:

$$E(\widehat{u}_1) = (i_H a_{1H} w_H + i_L a_{1L} w_L)\sigma_{u_1}$$

$$V(\widehat{u}_1) = \left[(1 + i_H x_H)a_{1H}^2 w_H + (1 + i_L x_L)a_{1L}^2 w_L\right]\sigma_{u_1}^2$$

$$E(\widehat{u}_2 = (i_H a_{1H} w_H + i_L a_{1L} w_L)a_2^2 r_G \sigma_{u_2}$$

$$V(\widehat{u}_2) = \left[(1 + a_{1H}^2 r_G a_2^2 i_H x_H)w_H + (1 + a_{1L}^2 r_G a_2^2 i_L x_L)w_L\right]a_2^2 \sigma_{u_2}^2$$

$$Cov(\widehat{u}_1\widehat{u}_2) = \left[a_{1H}^2(1 + i_H x_H)w_H + a_{1L}^2(1 + i_L x_L)\right]a_2^2 r_G \sigma_{u_1}\sigma_{u_2}.$$

The resulting $Corr(\widehat{u}_1\widehat{u}_2)$ reduces to text equation [4] only if $a_{1H} = a_{1L}$ and selection is symmetrical (ie, $w_H = w_L$ and $i_H = -i_L$).

ACKNOWLEDGMENT

REFERENCES

Becker WA (1984) *Manual of Quantitative Genetics*. Academic Enterprises, Pullman, WA, 4th edn

Blanchard JP, Everett RW, Searle SR (1983) Estimation of genetic trends and correlations for Jersey cattle. *J Dairy Sci* 66, 1947-1954

Calo LL, McDowell RE, VanVleck LD, Miller PD (1973) Genetic aspects of beef production among Holstein-Friesians pedigree selected for milk production. *J Anim Sci* 37, 676-682

DeNise RSK, Ray DE (1987) Postweaning weights and gains of cattle raised under range and gain test environments. *J Anim Sci* 64, 969-976

Diaz C (1992) The effect of selection on the covariance between direct and maternal breeding value predictions in different environments. Ph D Diss, Virginia Polytechnic Institute and State Univ, Blacksburg, Virginia

Dickerson GE (1962) Implications of genetic-environmental interaction in animal breeding. *Anim Prod* 4, 47-64

Dijkstra J, Oldenbroek JK, Korver S, VanderWerf JHJ (1990) Breeding for veal and beef production in Dutch Red and White cattle. *Livest Prod Sci* 25, 183-198

Do C (1991) Improvement in accuracy using records lacking sire information in the animal model. Ph D Diss, Virginia Polytechnic Institute and State Univ, Blacksburg, Virginia

Falconer DS (1952) The problem of environment and selection. *Am Nat* 86, 293-298

Greenhalgh SA, Quaas RL, VanVleck LD (1986) Approximating prediction error variances for multiple trait sire evaluations. *J Dairy Sci* 69, 2877-2883

Henderson CR (1963) *Selection Index And Expected Genetic Advance.* Natl Acad Sci Natl Res Council Publ 982, Washington, DC

Henderson CR (1973) Sire evaluation and genetic trands. *In: Proc Anim Breed Genet Symp in Honor of Dr Jay L Lush.* Am Soc Anim Sci, Champaign, IL, 10-41

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario

Hill WG (1970) Design of experiments to estimate heritability by regression of offspring on selected parents. *Biometrics* 26, 566-571

Hill WG, Thompson R (1977) Design of experiments to estimate offspring parent regression using selected parents. *Anim Prod* 24, 163-168

Johnson NL, Kotz S (1970) *Distributions in Statistics: Continuous Univariate Distributions-1.* John Wiley and Sons, New York

Mahrt GS, Notter DR, Beal WE, McClure WH, Bettison LG (1990) Growth of crossbred progeny of Polled Hereford sires divergently selected for yearling weight and maternal ability. *J Anim Sci* 68, 1889-1898

Mallinckrodt CH (1990) Relationship of milk expected progeny differences to actual milk production and calf weaning weight. M Sc thesis, Colorado State Univ, Fort Collins, CO

Meyer K (1989) Approximate accuracy of genetic evaluation under an animal model, *Livest Prod Sci* 21, 87-100

Meyer K (1991) Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genet Sel Evol* 23, 67-83

Oldenbroek JK, Meijering A (1986) Breeding for veal production in Black and White dairy cattle. *Livest Prod Sci* 15, 325-336

Robertson A (1959) The sampling variance of the genetic correlation coefficient. *Biometrics* 15, 469-485

Robertson A (1966) A mathematical model of the culling process in dairy cattle. *Anim Prod* 8, 95-108

Robertson A (1977) The effect of selection on the estimation of genetic parameters. *Z Tierzuchtg Zuchtgsbiol* 94, 131-135

Robinson GK, Jones LP (1987) Approximations for prediction error variances. *J Dairy Sci* 70, 1623-1632

SAS (1985) *SAS User's Guide: Statistics.* SAS Institute Inc, Cary, NC

Searle SR (1966) *Matrix Algebra for the Biological Sciences.* John Wiley and Sons, NY

Snedecor GW, Cochran WG (1967) *Statistical Methods.* The Iowa State Univ Press, Ames, IA, 6th edn

Taylor J (1983) Assumptions required to approximate unbiased estimates of genetic (co)variance by the method of Calo *et al* (1973). *In: Genetics Research 1982-1983.* Rep Eastern Artificial Insemination Coop, Dept Anim Sci Cornell Univ, Ithaca, NY, 256-261

Tilsch K, Wollert J, Bauming A (1989a) Relationships between breeding values for growth of beef sires in purebreeding and crossbreeding. *Livest Prod Sci* 21, 275-285

Tilsch K, Wollert J, Nurnberg G (1989b) Studies on sire x sex/environment interactions and their effect of response to selection in beef sires progeny-tested for fattening performance and carcass yield. *Livest Prod Sci* 21, 287-302

VanRaden PM, Freeman AE (1985) Rapid method to obtain bounds on accuracies and prediction error variances in mixed models. *J Dairy Sci* 68, 2123-2133

VanRaden PM, Wiggans GR (1991) Derivation, calculation, and use of animal model information. *J Dairy Sci*, 74, 2737-2746

Yamada Y (1962) Genotype by environment interaction and genetic correlation of the same trait under different environment. *Jpn J Genet* 37, 498-509