Proceedings, 10$^{th}$ World Congress of Genetics Applied to Livestock Production

# The Standard Error of the Estimated Purebred-Crossbred Genetic Correlation

**P. Bijma** and J. W. M. Bastiaansen

Animal Breeding and Genomics Centre, Wageningen University, Wageningen, the Netherlands

**ABSTRACT:** The purebred-crossbred genetic correlation ($r_{pc}$) is the key parameter determining the need for crossbred information. This work presents a simple equation for the standard error of the estimated $r_{pc}$ for nested full-half sib schemes with common-litter effects. The result shows that the standard error of $r_{pc}$ is determined by its true value, the number of sire families, and the accuracy of sire EBVs. Application to typical breeding schemes shows that the required number of sire families is large, usually exceeding 100, particularly when the numbers of dams per sire is small and common-litter effects are large. This work allows *a priori* optimization of data collection with the aim to estimate $r_{pc}$.

**Keywords:** crossbreeding; genotype by environment interaction; genetic correlation; standard error

## Introduction

The ultimate goal of crossbreeding schemes is to improve performance of the crossbred offspring of the pure breeding lines. Due to non-additive effects and genotype by environment interaction, the purebred performance is often an imperfect predictor of crossbred performance. In that case, selection in crossbreeding schemes should ideally be based on information recorded on crossbred individuals. The genetic correlation between the purebred and crossbred trait ($r_{pc}$) is the key parameter determining the need for crossbred information (Wei et al. (1991); Wei and Van der Werf (1994)). Hence, knowledge of $r_{pc}$ is required to decide on the strategy for data collection in crossbreeding schemes.

With the availability of genomic prediction methodology, interest in utilizing crossbred information is increasing (Ibáñẽz-Escriche et al. (2009)). While the use of crossbred phenotypes has been limited in traditional breeding programs, because tracing pedigree relationships in a crossbred production environment is non-trivial, it has regained attention recently because genomic relations are an alternative for the cumbersome pedigree tracing process. The idea that a "project" to build a training dataset with crossbred phenotypes will allow selection for crossbred performance is attractive and has revived interest in using crossbred phenotypes. The need for such a project increases when $r_{pc}$ differs more from one.

Prediction of the standard error (SE) of the estimated genetic correlation has considerably history (*e.g.,* Robertson (1959), Tallis (1959)). Particularly Robertson (1959) considered the standard error of the genetic correlation as a measure of genotype-by-environment interaction, of which $r_{pc}$ is a special case. Robertson (1959), however, considered schemes with equal heritability for both traits, and without full sibs nested within half sib families and common-litter effects. It is unclear, therefore,

whether previous results extend to relevant cases in, e.g., pigs, poultry or aquaculture.

Here we present a simple expression for the standard error of the estimated purebred-crossbred genetic correlation, which allows a priori optimization of designs aiming to estimate $r_{pc}$. An impression of required sample sizes is given for a number of practical cases.

## Material and Methods

Consider $N$ sires. Each sire is mated to $n_{d_p}$ dams of its own line, each dam producing $n_{o_p}$ purebred offspring, and to $n_{d_c}$ dams of the other line, each dam producing $n_{o_c}$ crossbred offspring. Throughout subscript $p$ denotes purebreds and $c$ crossbreds. Thus a half sib structure was assumed between purebreds and crossbreds, whereas full-sibs families were nested within half-sib families within purebreds and crossbreds.

For both traits, the model was

$$P_i = A_i + c_i + e_i,$$

$A$ denoting the breeding value, and $c$ the common litter effect. The estimated purebred-crossbred genetic correlation is given by

$$\hat{r}_{pc} = \frac{\hat{\sigma}_{A_{pc}}}{\hat{\sigma}_{A_p} \hat{\sigma}_{A_c}}.$$

The standard error of $\hat{r}_{pc}$ was approximated using a Taylor-series expansion, giving

$$SE(\hat{r}_{pc}) \approx \sqrt{\frac{\text{var}(\hat{\sigma}_{A_{pc}})}{\sigma_{A_p}^2 \sigma_{A_c}^2} + \frac{\sigma_{A_{pc}}^2 \text{var}(\hat{\sigma}_{A_p} \hat{\sigma}_{A_c})}{\sigma_{A_p}^4 \sigma_{A_c}^4} - \frac{2\sigma_{A_{pc}} \text{cov}(\hat{\sigma}_{A_{pc}}, \hat{\sigma}_{A_p} \hat{\sigma}_{A_c})}{\sigma_{A_p}^3 \sigma_{A_c}^3}}$$

Variances of components of $\hat{r}_{pc}$ were obtained from ANOVA within both populations, and from the cross-product of both progeny means of sires between populations. Inclusion of covariances between components of $\hat{r}_{pc}$ proved essential, and required terms were obtained using Taylor-series expansions and the law of total variance (derivations not shown). The resulting expression was accurate but complex.

A considerable simplification was achieved by defining an effective number of sire families, which depended on the reliability of sire EBVs

$$N_{eff} = \rho^4 (N-1),$$

in which $\rho^4$ is the squared reliability of sire-EBVs,

$$\rho^2 = \frac{\tfrac{1}{4}\sigma_A^2}{\mathrm{var}(\overline{P})},$$

$$\mathrm{var}(\overline{P}) = \tfrac{1}{4}\sigma_A^2 + \frac{\tfrac{1}{4}\sigma_A^2 + \sigma_c^2}{n_d} + \frac{\tfrac{1}{2}\sigma_A^2 + \sigma_e^2}{n_d n_o},$$

the $\overline{P}$ denoting the progeny-mean of a sire. With this substitution, the variances of the estimated purebred and crossbred genetic variances became simple functions of reliability,

$$\mathrm{var}(\hat{\sigma}_A^2) \approx \frac{2\sigma_A^4}{\rho^2(N-1)},$$

which allowed considerable simplification. Finally, a simple expression for the standard error of $\hat{r}_{pc}$ was obtained,

$$SE(\hat{r}_{pc}) \approx \sqrt{\frac{\frac{1}{\rho_p^2 \rho_c^2} + \left(1 + \frac{0.5}{\rho_p^4} + \frac{0.5}{\rho_c^4} - \frac{2}{\rho_p^2} - \frac{2}{\rho_c^2}\right) r_{pc}^2 + r_{pc}^4}{N-1}}.$$

This result shows that the SE of $\hat{r}_{pc}$ is a simple function of the true value of $r_{pc}$ and of the reliabilities of sire-EBVs for purebred and crossbred performance. Stochastic simulations of 320 alternative scenarios with 1000 replicates each showed that the expression is accurate, with an average absolute relative error of 3.1%.

## Results

Figure 1 shows $SE(\hat{r}_{pc})$ as a function of $r_{pc}$ for a sample size of 100 sires, and for different reliabilities of sire EBVs. When sire EBVs are accurate, $SE(\hat{r}_{pc})$ is considerably smaller when $r_{pc}$ is close to 1. When sire EBVs are inaccurate, however, there is only a weak relationship between $SE(\hat{r}_{pc})$ and $r_{pc}$. Clearly, a sample of 100 half-sib families is too small, unless reliabilities of sire EBVs are close to one and $r_{pc} > {\sim}0.7$.

Figure 2 shows $SE(\hat{r}_{pc})$ as a function of the number of half-sib families, for a range of schemes that represent practical cases (personal communication Egiel Hanenberg, Gosse Veninga, Hooi Ling Khaw and Jeroen Visscher). Results show that the common strategy in some aquaculture breeding programs such as Tilapia to mate a sire to only two dams, together with the presence of common full-sib family effects, causes very large standard errors, even when 600 half sib families are used. On the other hand, the use of large numbers of dams per sire in broiler breeding causes standard errors to approach their minimum possible value.

**Figure 1.** $SE(\hat{r}_{pc})$ as a function of $r_{pc}$, for different reliabilities of sire EBVs (r2) that are assumed to be the same for the purebred and crossbred trait. For $N = 100$. The figure is symmetric in $r_{pc}$, so the range for $r_{pc} = -1$ to 0 is omitted.
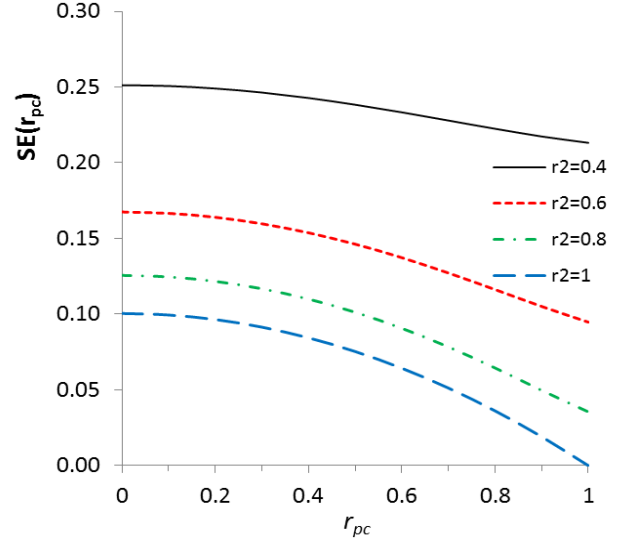


## Discussion

This work has presented a simple expression for the standard error of the estimated purebred-crossbred genetic correlation. Based on Robertson's (1959) results, Falconer and Mackay (1996) presented a simplified prediction for the SE of an ordinary genetic correlation, taking the form $SE(\hat{r}_g) = (1 - r_g^2)x$, where $x$ is a function of the data structure and heritabilities. When applied to the purebred-crossbred genetic correlation, this expression yields $SE(\hat{r}_{pc}) = 0$ at $r_{pc} = 1$, which is very inaccurate unless reliabilities of sire EBVs are one (Figure 1). For $r_{pc} \to 1$ and equal accuracies of sire EBVs for both traits, the above expression reduces to

$$SE(\hat{r}_{pc} \mid r_{pc} = 1) \approx \frac{\sqrt{2}}{\sqrt{N-1}}\left(\frac{1}{\rho^2} - 1\right)$$

which does not approach zero unless reliabilities approach one (See values for $r_{pc} = 1$ in Figure 1).

## Conclusion

This work has presented a simple and accurate prediction of the standard error of the estimated purebred-crossbred genetic correlation. This allows breeders to decide on the required sample size to estimate this correlation, so as to support decisions on the collection of crossbred information. Results show that >100 half sib families will be required in most cases.

**Figure 2.** $SE(\hat{r}_{pc})$ examples for typical breeding schemes as a function of the number of half-sib families (*N*). Input values: Harvest weight in aquaculture: $n_d = 2$, $n_o = 40$, $h^2 = 0.3$, $c^2 = 0.15$, $r_{pc} = 0.8$. Egg number in laying hens: $n_d = 7$, $n_{o_p} = 10$, $n_{o_c} = 5$, $h^2 = 0.2$, $c^2 = 0$, $r_{pc} = 0.6$. Growth rate in pigs: $n_d = 10$, $n_o = 10$, $h^2 = 0.3$, $c^2 = 0.10$, $r_{pc} = 0.7$. Growth rate in broilers: $n_d = 12$, $n_{o_p} = 70$, $n_{o_c} = 10$, $h^2 = 0.3$, $c^2 = 0.05$, $r_{pc} = 0.8$. Minimum: Lowest possible $SE(\hat{r}_{pc})$ for $r_{pc} = 0.7$; refers to a scheme with reliability of sires EBVs equal to 1.
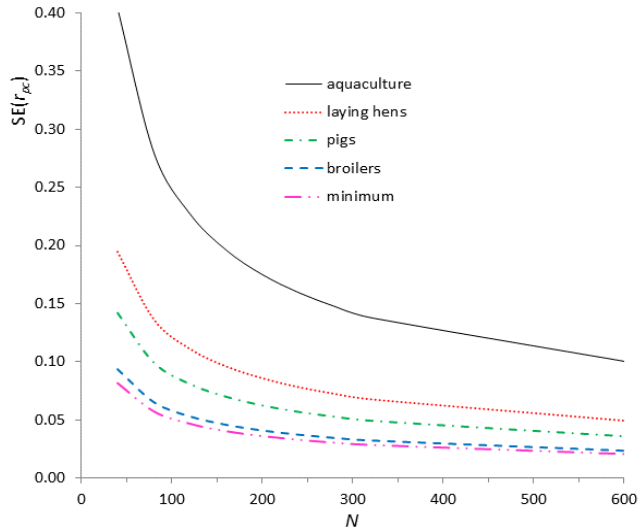
## Literature Cited

Falconer, D.S., Mackay, T.F.C. (1996) *Introduction to quantitative genetics.* Longman, Essex, UK.

Ibáñẽz-Escriche, N., Fernando, R.L., Toosi, A. et al. (2009). Genet. Sel. Evol. 41:12.

Robertson, A. (1959). Biometrics 15:469-485.

Tallis, G.M. (1959). Aust. J. Stat. 1:35-43.

Wei, M., Van der Werf, J. H. J. (1994). Anim. Prod. 59:401-413.

Wei M, Van der Steen H. A. M., Van der Werf, J. H. J. et al. (1991). J. Anim. Breed. Genet. 108:253-261.