

RESEARCH

Open Access

# Standard error of the genetic correlation: how much data do we need to estimate a purebred-crossbred genetic correlation?

Piter Bijma\* and John WM Bastiaansen

## Abstract

**Background:** The additive genetic correlation ( $r_g$ ) is a key parameter in livestock genetic improvement. The standard error (SE) of an estimate of  $r_g$ ,  $\hat{r}_g$ , depends on whether both traits are recorded on the same individual or on distinct individuals. The genetic correlation between traits recorded on distinct individuals is relevant as a measure of, e.g., genotype-by-environment interaction and for traits expressed in purebreds vs. crossbreds. In crossbreeding schemes,  $r_g$  between the purebred and crossbred trait is the key parameter that determines the need for crossbred information. This work presents a simple equation to predict the SE of  $\hat{r}_g$  between traits recorded on distinct individuals for nested full-half sib schemes with common-litter effects, using the purebred-crossbred genetic correlation as an example. The resulting expression allows *a priori* optimization of designs that aim at estimating  $r_g$ . An R-script that implements the expression is included.

**Results:** The SE of  $\hat{r}_g$  is determined by the true value of  $r_g$ , the number of sire families ( $N$ ), and the reliabilities of sire estimated breeding values (EBV):

$$SE(\hat{r}_g) \approx \sqrt{\frac{\frac{1}{\rho_x^2 \rho_y^2} + \left(1 + \frac{0.5}{\rho_x^4} + \frac{0.5}{\rho_y^4} - \frac{2}{\rho_x^2} - \frac{2}{\rho_y^2}\right) r_g^2 + r_g^4}{N-1}}$$

where  $\rho_x^2$  and  $\rho_y^2$  are the reliabilities of the sire EBV for both traits. Results from stochastic simulation show that this equation is accurate since the average absolute error of the prediction across 320 alternative breeding schemes was 3.2%. Application to typical crossbreeding schemes shows that a large number of sire families is required, usually more than 100. Since  $SE(\hat{r}_g)$  is a function of reliabilities of EBV, the result probably extends to other cases such as repeated records, but this was not validated by simulation.

**Conclusions:** This work provides an accurate tool to determine *a priori* the amount of data required to estimate a genetic correlation between traits measured on distinct individuals, such as the purebred-crossbred genetic correlation.

\* Correspondence: piter.bijma@wur.nl  
Animal Breeding and Genomics Centre, Wageningen University, PO Box 338,  
6700 AH Wageningen, The Netherlands

## Background

The additive genetic correlation is a key parameter in livestock genetic improvement and is defined as the correlation between breeding values of individuals for two distinct traits, say  $x$  and  $y$  [1],

$$r_g = \frac{\sigma_{A_{xy}}}{\sigma_{A_x} \sigma_{A_y}},$$

where  $\sigma_{A_{xy}}$  denotes the covariance between the breeding values  $A_x$  and  $A_y$  of individuals, and  $\sigma_{A_x}$  and  $\sigma_{A_y}$  the additive genetic standard deviations. Estimation of  $r_g$  requires substantial amounts of data [2-4].

The standard error (SE) of the estimated genetic correlation depends on whether both traits are recorded on the same individual or on distinct individuals [2]. Examples of cases where both traits are recorded on distinct individuals are: (i) traits that are expressed in different environments, where  $r_g$  is a measure of the degree of genotype-by-environment interaction, (ii) traits that are expressed in males *vs.* females, such as sperm quality in bulls and milk yield in cows, (iii) traits that are expressed in live *vs.* dead animals, such as meat quality traits in fattening pigs and longevity of sows, and (iv) traits that are expressed in purebreds *vs.* crossbreds. This work considers the SE of the estimated genetic correlation between traits recorded on distinct individuals, with a focus on the purebred-crossbred genetic correlation.

In crossbreeding schemes, the ultimate goal is to improve the performance of the crossbred offspring of the pure breeding lines. With genotype-by-environment interaction and/or non-additive genetic effects, purebred performance is an imperfect predictor of crossbred performance. Thus, selection in crossbreeding schemes is ideally based on information recorded on crossbred relatives of the purebred selection candidates, or on a genomic reference population based on crossbred phenotypes [5-8]. However, phenotypic and pedigree data are not always routinely collected on crossbred individuals. The genetic correlation between the purebred and the crossbred trait ( $r_{pc}$ ) is the key parameter that determines the need for crossbred information. Hence, accurate estimation of  $r_{pc}$  [9,10] is required to decide on the strategy used for data recording.

A priori, the desired accuracy of an estimate of  $r_{pc}$  should be at least as high as for an ordinary genetic correlation. For example, when accuracies of purebred and crossbred EBV (estimated breeding values) are similar, the loss in response to selection due to relying on purebred rather than crossbred information is  $\sim 10\%$  when  $r_{pc}$  is 0.9, but  $\sim 30\%$  when  $r_{pc}$  is 0.7. To accurately identify such differences in  $r_{pc}$  the SE of the estimated correlation should not be greater than  $\sim 0.05$ .

Predicting the SE of estimates of the genetic correlation has been studied for many years [2-4,11]. In particular, Robertson [2] considered the SE of estimates of the genetic correlation between traits recorded on distinct

individuals, such as  $r_{pc}$  [12], but only for cases with equal heritabilities and equal numbers of offspring for both traits. Moreover, the reports in [2-4,11] all considered half-sib designs, and did not allow for full-sib groups within half-sib families or for common-litter environmental effects.

In addition, existing prediction equations may not be readily accessible to applied breeders, because the full predictions are complex and expressed in terms of intra-class correlations, rather than heritabilities and common-litter variances. Simplified expressions do exist, but express the SE as being proportional to  $\left(1 - r_g^2\right)$  and are very inaccurate when  $r_g$  is close to 1, which may often be the case for a genotype-by-environment correlation or purebred-crossbred correlation [1,2,4]. With the computing power available today, stochastic simulations offer a solution, but they are still too time-consuming to use as a simple interactive tool. Thus, although the topic is somewhat outdated, for applied breeding it is still relevant to propose a simple prediction of the SE of estimates of genetic correlations.

Moreover, while the use of crossbred phenotypes has been limited in applied breeding programs because tracing pedigree relationships in a crossbred production environment is not trivial, it has recently regained attention because genomic relations are a solution for the cumbersome pedigree tracing process. The idea that building a training dataset with crossbred phenotypes will permit selection for crossbred performance is attractive and has revived interest in using crossbred phenotypes.

Here, we present a simple prediction equation for the SE of the estimated genetic correlation between traits recorded on distinct individuals, for nested full-half sib schemes with common-litter effects. This expression allows *a priori* optimization of designs that aim at estimating  $r_g$ . To facilitate application, an R-script that implements the prediction is included in Additional file 1. Examples of sample sizes required to estimate  $r_{pc}$  are provided for a number of practical cases, but optimization of schemes is not considered extensively, since it can be easily done for specific cases using the R-script.

## Methods

### Analytical prediction of the SE of genetic correlation estimates

In the following, purebred and crossbred performance will be used as an example of two traits recorded on distinct individuals. Hence, subscript  $p$ , referring to purebred, will be used to denote one trait, and subscript  $c$ , referring to crossbred, to denote the other. However, the resulting expression will apply to the general case of a genetic correlation between traits recorded on distinct individuals.

Consider a population with phenotypic records on purebred and crossbred offspring of  $N$  sires. Each sire was 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. Thus, a half-sib structure is present between purebreds and crossbreds, whereas full-sib families are nested within half-sib families within the purebreds and within the crossbreds.

For both purebreds and crossbreds, the trait model is given by:

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

where  $A_i$  denotes the breeding value,  $c_i$  the common-litter effect, and  $e_i$  the environmental effect for trait  $i$  (purebred or crossbred). Hence, it is assumed implicitly that fixed effects can be estimated accurately. We do not model permanent environmental effects. Hence, a single observation per individual and a single litter per dam are assumed.

The estimate of the purebred-crossbred genetic correlation is given by:

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

where  $\hat{\sigma}_{A_{pc}}$  denotes the estimate of the purebred-crossbred genetic covariance, and  $\hat{\sigma}_{A_p}$  and  $\hat{\sigma}_{A_c}$  the estimates of genetic standard deviations. Throughout this article, symbols with hats (^) denote estimates, which are random variables, while symbols without hats denote the true parameters. The standard error of  $\hat{r}_{pc}$  was derived using a Taylor-series expansion of the expression for  $\hat{r}_{pc}$ . The final result is presented in the main text, while derivations are in Additional file 2.

The resulting expression shows that the SE of the estimate of the purebred-crossbred genetic correlation is determined by the true value of  $r_{pc}$ , the number of sire families,  $N$ , and the reliabilities of sire EBV,

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

where  $\rho_p^2$  is the reliability (*i.e.*, squared accuracy) of sire EBV for purebred performance, and  $\rho_c^2$  the reliability of sire EBV for crossbred performance. Reliabilities of EBV are given by:

$$\rho^2 = \frac{1/4 \sigma_A^2}{\text{var}(\bar{P})}, \quad (2)$$

where  $\bar{P}$  denotes the average phenotypic value of the progeny of a sire with a variance equal to:

$$\text{var}(\bar{P}) = 1/4 \sigma_A^2 + \frac{1/4 \sigma_A^2 + \sigma_c^2}{n_d} + \frac{1/2 \sigma_A^2 + \sigma_e^2}{n_d n_o}, \quad (3)$$

where  $\sigma_c^2$  denotes the common-litter variance and  $\sigma_e^2$  the environmental variance. Thus, Equations 2 and 3 are

used twice, once for purebreds and once for crossbreds. Instead of using Equations 2 and 3, empirical reliabilities from genetic evaluations, when available, can be substituted into Equation 1.

In the limiting case where the number of dams mated to a sire and the number of offspring per dam are large, so that  $\rho_p^2 = \rho_c^2 \rightarrow 1$ , the expression reduces to:

$$SE(\hat{r}_{pc}) \approx \frac{1-r_{pc}^2}{\sqrt{N-1}}, \quad (4)$$

which is the common expression for the SE of a simple correlation coefficient [13].

### Simulations

A limited number of scenarios was tested by estimation of  $r_{pc}$  in simulated data using ReML [14] and compared to results from analysis of the data using random-effects ANOVA with dam families nested within sire families [15] and to predictions from Equation 1. The simulated data consisted of sires with purebred and crossbred offspring. Crossbred offspring were from F1 females mated to a terminal sire line, *i.e.*, three purebred lines were simulated, each with an  $N_e$  of 100. For each purebred line, 10 generations of pedigree were used. Purebred and crossbred phenotypes were simulated from multivariate normal distributions, for different values of  $h_p^2$ ,  $h_c^2$ , and  $r_{pc}$ . Genetic correlations were estimated with the ASReml software [16], using 200 replicates per scenario. Average  $SE(\hat{r}_{pc})$  as reported by ASReml and the standard deviation of  $\hat{r}_{pc}$  over the 200 replicates were calculated.

A large number of simulated scenarios was tested using ANOVA and compared to predictions from Equation 1. One thousand replicates of all factorial combinations of  $N = (50, 150)$ ,  $n_{dp} = 10$ ,  $n_{d_c} = (5, 20)$ ,  $n_{o_p} = 8$ ,  $n_{o_c} = (6, 12)$ ,  $r_{pc} = (-0.8, -0.4, 0, 0.4, 0.8)$ ,  $h_p^2 = (0.3, 0.6)$ ,  $h_c^2 = (0.2, 0.4)$ ,  $c_p^2 = 0.05$  and  $c_c^2 = (0, 0.1)$  were simulated (320 scenarios in total). Genetic parameters were estimated using ANOVA. Estimates of  $r_{pc}$  outside the boundaries of  $-1$  and  $1$  were set to the nearest boundary.

### Results

#### Accuracy of SE predictions

Concordance between the ReML and ANOVA estimates from the simulations was very high (Table 1). The SE from the ReML analyses were a little lower than the SE from the ANOVA estimates, which was expected because the ReML estimates used 10 generations of pedigree information, whereas the ANOVA estimates were based on a family structure of a single generation. Moreover, the SE of the ReML estimates were less precisely estimated because of the limited number of replicates (See footnote of Table 1). Because of computation time,

**Table 1 Comparison of predicted  $SE(\hat{r}_{pc})$  from Equation 1 to empirical estimates from ANOVA and to empirical and reported estimates from ASReml**

Design						SE( $\hat{r}_{pc}$ )			
$r_{pc}$	$h_p^2$	$h_c^2$	$n_d$	$n_o$	$N$	Equation 1	<sup>1</sup> Anova	<sup>2</sup> ASReml Reported	<sup>3</sup> ASReml Empirical
0.4	0.1	0.1	10	4	100	0.195	0.203	0.191	<sup>4</sup> 0.250
0.8	0.5	0.5	10	4	100	0.065	0.066	0.061	0.060
0.0	0.3	0.3	20	4	100	0.120	0.123	0.118	0.127
0.4	0.3	0.3	20	4	100	0.104	0.104	0.103	0.105
0.0	0.1	0.1	10	4	200	0.145	0.146	0.143	0.146
-0.8	0.5	0.5	10	8	200	0.039	0.039	0.036	0.034
0.8	0.5	0.5	20	8	200	0.032	0.032	0.028	0.030

For  $\sigma_c^2 = 0$ ; <sup>1</sup>results are the SD among 1000 replicates of  $\hat{r}_{pc}$ ; <sup>2</sup>results are the average of reported SE of 200 replicates; <sup>3</sup>results are the SD among 200 replicates of  $\hat{r}_{pc}$ ; <sup>4</sup>four replicates were fixed at the boundary of  $\hat{r}_{pc}$  1; with these four estimates removed the SE equaled 0.216; <sup>5</sup>Empirical SE from ASReml were based on 200 replicates only, and may therefore deviate from the true SE. With 200 replicates, the SE of the relative empirical SE, i.e. the SE of the ratio of the empirical SE over the true SE, equals  $SE[SE(\hat{r}_{pc})/SE(\hat{r}_{pc})] = 1/\sqrt{2 \times (200-1)} \approx 0.05$  [13]; thus a 5% error in predicted SE does not indicate a significant discrepancy between predictions and simulations, indicating that 200 replicates yield a limited accuracy of the empirical SE; when predicted  $SE(\hat{r}_{pc})$  is unbiased, the expected absolute relative error equals  $\approx 3.5\%$ , and a relative error  $>9.8\%$  indicates a significant difference between empirical and predicted  $SE(\hat{r}_{pc})$  ( $P < 0.05$ ; two-sided, not accounting for multiple testing).

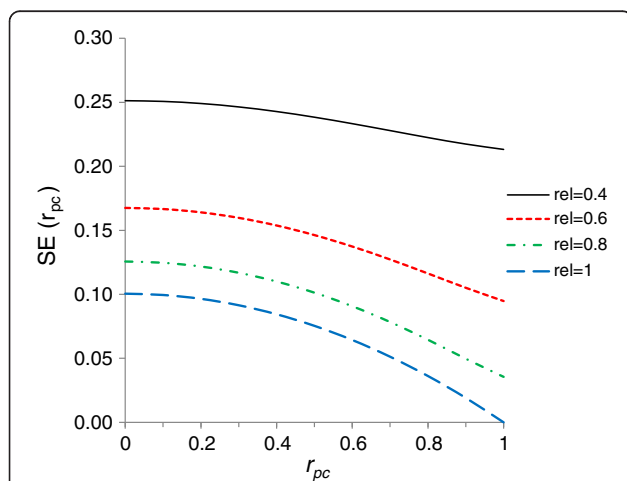
more extensive evaluation of the accuracy of predictions from Equation 1 was based on the ANOVA estimates.

ANOVA estimates showed that the predicted SE from Equation 1 were accurate since the average absolute relative error across all schemes evaluated was equal to 3.2% (=100%  $\times$  |predicted SE-simulated SE|/simulated SE; [see Additional file 3]). Sizeable errors occurred only for schemes for which estimates of genetic variances were near 0 in some replicates, which yielded extreme values for  $\hat{r}_{pc}$  (this occurred occasionally for schemes with  $N = 50$ ,  $h_c^2 = 0.2$  and  $c_c^2 = 0.1$ ). For those schemes, the maximum absolute relative error was 14%. These

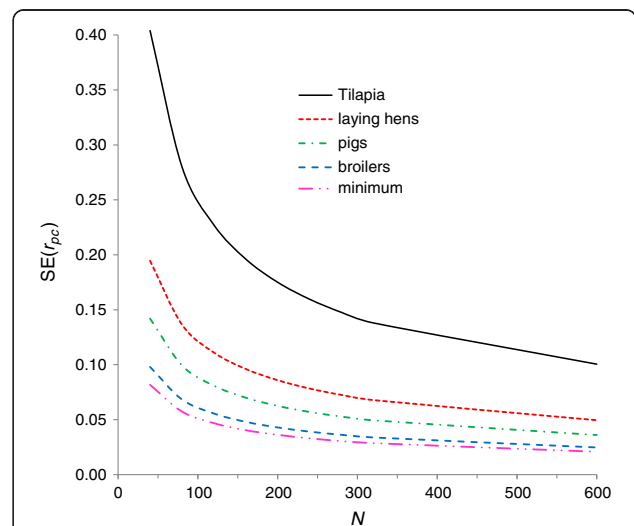
schemes are, however, of little practical relevance since their  $SE(\hat{r}_{pc})$  was around 0.25, which is far too high to be useful in practice.

#### Required sample sizes

Figure 1 shows predictions of  $SE(\hat{r}_{pc})$  based on Equation 1 as a function of  $r_{pc}$  for a sample size of 100 sires, and for



**Figure 1 Predictions of  $SE(\hat{r}_{pc})$  as a function of  $r_{pc}$  for different reliabilities of sire EBV (rel) 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.  $SE(\hat{r}_{pc})$  was based on Equation 1.



**Figure 2 Predictions of  $SE(\hat{r}_{pc})$  for typical breeding schemes as a function of the number of half-sib families ( $N$ ).**  $SE(\hat{r}_{pc})$  were based on Equation 1 with the following input values: Harvest weight in aquaculture (tilapia):  $n_d = 2$ ,  $n_o = 40$ ,  $h^2 = 0.3$ ,  $c^2 = 0.15$ ,  $r_{pc} = 0.8$ . Egg number in laying hens:  $n_d = 7$ ,  $n_o = 10$ ,  $n_{oc} = 5$ ,  $h^2 = 0.2$ ,  $c^2 = 0$ ,  $r_{pc} = 0.6$ . 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:  $n_d = 12$ ,  $n_{op} = 70$ ,  $n_{oc} = 10$ ,  $h^2 = 0.3$ ,  $c^2 = 0.05$ ,  $r_{pc} = 0.8$ . Minimum: Lowest possible  $SE(\hat{r}_{pc})$  for  $r_{pc} = 0.7$  refers to a scheme with many dams per sire and many offspring per dam.

different reliabilities of sire EBV. When sire EBV have high reliability,  $SE(\hat{r}_{pc})$  becomes considerably smaller when  $r_{pc}$  comes closer to 1. However, when sire EBV are inaccurate 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 EBV are close to 1 and  $r_{pc}$  is greater than  $\sim 0.7$ .

Figure 2 shows predictions of  $SE(\hat{r}_{pc})$  as a function of the number of half-sib families, for a range of schemes that may represent practical cases (personal communication Egiel Hanenberg, Gosse Veninga, Hooi Ling Khaw and Jeroen Visscher). Results from aquaculture breeding programs, such as for Tilapia, show that the commonly used strategy of mating 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 contrary, the use of large numbers of dams per sire in broiler chicken breeding causes standard errors to approach their theoretical minimum (Equation 4).

## Discussion

The main objective of this work was to provide breeders with a simple tool to predict the SE of estimates of the genetic correlation between traits recorded on distinct individuals ( $SE(\hat{r}_{pc})$ ). The objective was not to address theoretical issues underlying the SE of genetic correlation estimates, which have been discussed extensively in the past [2-4,11]. Nevertheless, this work provides new insight on the impact of the reliability of sire EBV on  $SE(\hat{r}_{pc})$ , which was not obvious from previous work. Equation 1 shows that  $SE(\hat{r}_{pc})$  depends on the reliabilities of sire EBV and the true value of  $r_{pc}$ . Since Equation 1 is expressed in terms of reliabilities, it probably extends to other models for trait analysis, such as repeatability models, but this was not validated by simulation.

On the one hand, Equation 1 can be interpreted as a lower bound of  $SE(\hat{r}_{pc})$  because it assumes a balanced design and that the fixed effects are known, while actual estimation of  $r_{pc}$  always involves somewhat unbalanced data and estimation of fixed effects. However, on the other hand, Equation 1 assumes that  $r_{pc}$  is estimated from half-sib relationships only, whereas estimation of genetic parameters in livestock populations usually includes multiple generations of pedigree information, so that more distant relationships also contribute to the estimate, which reduces the SE.

We have considered a genetic correlation between traits measured on distinct individuals, of which the genetic correlation between purebred and crossbred performance,  $r_{pc}$ , is an important example. When both traits are measured on the same individuals, additional complications arise due to covariances between the dam,

common-litter and residual effects for the two traits. In such a case, derivation of  $SE(\hat{r}_g)$  for a nested full-half sib scheme with common-litter effects is complicated, and this was not attempted here. When both traits are measured on the same individuals, stochastic simulation results (not shown) indicate that  $SE(\hat{r}_g)$  is similar to the value given by Equation 1 when  $r_g = 0$ , but smaller than that value when the true correlation differs from 0. Hence, in most cases, the SE of a genetic correlation between traits measured on the same individuals is smaller than the value obtained from Equation 1.

Based on Robertson's results [2], Falconer and Mackay [1] presented a simplified prediction of  $SE(\hat{r}_g)$ , taking the form  $SE(\hat{r}_g) = (1-r_g^2)x$ , where  $x$  is a function of the data structure and heritabilities. For  $r_g = \pm 1$ , this expression yields  $SE(\hat{r}_g) = 0$ , which is very inaccurate unless the reliabilities of sire EBV are close to 1 (Figure 1). For  $r_g \rightarrow 1$  and equal reliabilities of sire EBV for both traits, Equation 1 reduces to:

$$SE(\hat{r}_g | r_g = 1) \approx \frac{\sqrt{2}}{\sqrt{N-1}} \left( \frac{1}{\rho^2} - 1 \right), \quad (5)$$

which does not approach 0 unless reliabilities approach 1 (see values for  $r_{pc} = 1$  in Figure 1).

## Conclusions

This paper presents a simple and accurate prediction of the standard error of estimates of the genetic correlation between traits recorded on distinct individuals, for nested full-half sibs schemes with common-litter effects. This allows breeders to decide on the required sample size to estimate this correlation, e.g., to support decisions on the collection of crossbred information. Results show that more than 100 half sib families are required in most cases.

## Additional files

**Additional file 1: R-code for SE(rpc).** This file contains an R-script that implements Equation 1 for a range of input values of genetic parameters and breeding designs.

**Additional file 2: Derivation of Equation 1.** This file contains the derivation of Equation 1.

**Additional file 3: Numerical validation of Equation 1.** This file shows a comparison of predicted and empirical SE for a range of alternative schemes.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

PB and JB together conceived the study. PB derived the mathematical results and drafted the initial manuscript. Both authors contributed to the stochastic simulations and the writing of the manuscript. All authors read and approved the final manuscript.

### Acknowledgements

We thank Mario Calus for discussion on this topic, and Egiel Hanenberg (Topigs Norsvin), Gosse Veninga (Cobb Europe), Jeroen Visscher (Hendrix-Genetics ISA) and Hooi Ling Khaw (World Fish Centre) for providing information on breeding schemes. The contribution of PB was supported by the foundation for applied sciences (STW) of the Dutch science council (NWO). The contribution of JB was supported by PPP Breed4Food.

Received: 10 January 2014 Accepted: 24 September 2014

Published online: 19 November 2014

### References

1. Falconer DS, Mackay TFC: *Introduction to Quantitative Genetics*. Essex: Longman Scientific and Technical; 1996.
2. Robertson A: **The sampling variance of the genetic correlation coefficient.** *Biometrics* 1959, **15**:469–485.
3. Tallis GM: **Sampling errors of genetic correlation coefficients calculated from analyses of variance and covariance.** *Aust J Stat* 1959, **1**:35–43.
4. Visscher PM: **On the sampling variance of intraclass correlations and genetic correlations.** *Genetics* 1998, **149**:1605–1614.
5. Wei M, Van der Werf JHJ: **Maximizing genetic response in crossbreds using both purebred and crossbred information.** *Anim Prod* 1994, **59**:401–413.
6. Bijma P, Van Arendonk JAM: **Maximizing genetic gain for the sire line of a crossbreeding scheme utilizing both purebred and crossbred information.** *Anim Sci* 1998, **66**:529–542.
7. Dekkers JCM: **Marker-assisted selection for commercial crossbred performance.** *J Anim Sci* 2007, **85**:2104–2114.
8. Ibánñez-Escriche N, Fernando RL, Toosi A, Dekkers JCM: **Genomic selection of purebreds for crossbred performance.** *Genet Sel Evol* 2009, **41**:12.
9. Wei M, Van der Werf JHJ: **Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits.** *J Anim Sci* 1995, **73**:2220–2226.
10. Lutaaya E, Misztal I, Mabry JW, Short T, Timm HH, Holzbauer R: **Genetic parameter estimates from joint evaluation of purebreds and crossbreds in swine using the crossbred model.** *J Anim Sci* 2001, **79**:3002–3007.
11. Reeve ECR: **The variance of the genetic correlation coefficient.** *Biometrics* 1955, **11**:357–374.
12. Wei M, Van der Steen HAM, Van der Werf JHJ, Brascamp EW: **Relationship between purebred and crossbred parameters.** *J Anim Breed Genet* 1991, **108**:253–261.
13. Stuart A, Ord JK: *Kendall's Advanced Theory of Statistics, Distribution theory Vol. 1*. London: Hodder Education; 1994.
14. Patterson HD, Thompson R: **Recovery of inter-block information when block sizes are unequal.** *Biometrika* 1971, **58**:545–554.
15. Stuart A, Ord JK, Arnold S: *Kendall's Advanced Theory of Statistics, Classical Inference and the Linear Model Vol. 2A*. 6th edition. London: Arnold; 1999.
16. Gilmour AR, Gogel BJ, Cullis BR, Thompson R: *ASReml User Guide Release 3.0*. Hemel Hempstead: VSN International Ltd; 2009.

doi:10.1186/s12711-014-0079-z

**Cite this article as:** Bijma and Bastiaansen: Standard error of the genetic correlation: how much data do we need to estimate a purebred-crossbred genetic correlation? *Genetics Selection Evolution* 2014 **46**:79.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

