Reliability of genetic evaluation using purebred and crossbred data from different pig farms

Dubravko Skorput[1]*, Gregor Gorjanc[2,3], Zoran Lukovic[1]

[1] Department of Animal Science and Technology, Faculty of Agriculture,
    University of Zagreb, Svetošimunska cesta 25, 10 000 Zagreb, Croatia

[2] Department of Animal Science, Biotechnical Faculty,
    University of Ljubljana, Groblje 3, 1230 Domžale, Slovenia

[3] The Roslin Institute, The University of Edinburgh,
    Easter Bush Midlothian, EH25 9RG Scotland, UK

The impact of connectedness between contemporary groups (CG) on the reliability of genetic evaluation was analysed in Landrace (L) and Large-White (LW) pigs and their crossbreds on 23 245 data records from three large scale farms and 70 family farms in the period of 2002-2012. Gene flow (GF) and connectedness rating (CR) methods were used to assess the information concerning the level of connectedness. The average reliability of estimated breeding values (EBV's) was calculated for all CGs in single and joint analyses for different combinations of data. Backfat thickness (BF) was the focal trait of the analysis. The average GF in the ten-year period was low and ranged from 0% of the genes originating from other farms on several CGs and 7% genes on farm B originating from farm C in L and from 0% to 6% of the genes originating from farm A on the group of family farms with respect to LW. The average CR was lower than 0.03 between all CGs. Thus, an unreliable comparison of estimated breeding values between animals from different herds could be expected. After inclusion of crossbred data, reliability increased from 0.08 to 0.27 for family farms in L and from 0.03 to 0.35 in the group of farms C in LW. Results showed the necessity to increase connectedness and consequently individual reliability of genetic evaluation and reliability of the across-herd comparison. Proper organisation of artificial insemination is proposed as an adequate tool to increase connectedness.

KEY WORDS: connectedness / crossbreeding / pigs / reliability

---

*Corresponding authot: dskorput@agr.hr

The BLUP procedure has been the method of choice for the genetic evaluation of pigs in the last few decades in conventional breeding programs. The procedure of genetic evaluation based on the mixed model methodology was introduced by Henderson [1949]. Along with the genetic evaluation of purebred animals, genetic merit is considered also for crossbreeds, due to their importance in achievement of economic benefits and their large number in commercial herds. The reliability of comparisons of breeding values and consequently selection decisions across contemporary groups (CG) are influenced by the degree of connectedness between those groups [Kennedy and Trus 1993, Mathur *et al.* 2002]. If the amount of genetic links between CGs in the genetic evaluation is insufficient, comparisons of breeding values between animals from different CGs may be biased. Such a problem was noticed in the species, in which the level of artificial insemination (AI) is low, such as beef cattle [Veselá *et al.* 2007], sheep [Lewis *et al.* 2005] as well as some pig breeding programs [Sun *et al.* 2009].

Comparisons among animals that are poorly connected will have a higher prediction error than those that are well connected. Decrease of variance for estimated differences between herds enables accurate ranking of animals across CGs [Kennedy and Trus 1993]. It is possible that the bias in accuracy of estimated breeding values (EBV) across herds in the case of low connectedness also has an unfavourable effect on the reliability of estimation for each breeding value from data sets, which contain animals from several herds having different genetic means [Smith and Banos 1991]. Several methods have been developed to measure the level of connectedness between CGs of livestock. Kennedy and Trus [1993] suggested that the most appropriate measure of connectedness would be the average prediction variance of differences in estimated breeding values between animals in different CGs. However, due to the high computational requirements for obtaining the prediction variance of differences in estimated breeding values, they proposed three alternative methods to describe the level of connectedness: gene flow (GF), genetic drift variance, and variance of estimates of CG effects. Mathur *et al.* [1998] expanded the latter method of Kennedy and Trus [1993] to the so-called connectedness rating (CR), expressed as the correlation between the estimates of CG effects.

The aim of this study was to determine the level of connectedness between CGs within a conventional breeding program based on purebred and crossbred data and to analyse the reliability of genetic evaluation-using different data sets.

Material and methods

Data records from the field test of young boars and gilts from the period between 2002 and 2012 were provided by the Croatian Agricultural Agency. The complete data set contained 23, 245 data records for four breeds: Landrace (L), Large White (LW), and their crossbreds, LxLW and LWxL on three large scale farms, for this purpose denoted as A, B, and C, and a group of 70 small farms, known as family farms, denoted as F. Family farms are characterised by a variable number of sows and different levels

of technological and management solutions among farms. Backfat thickness (BF) was the focal trait of the analysis. Data were removed when the animal participated in the field test for fewer than 120 and more than 360 days; animals weighed less than 70 kg and more than 120 kg; the herd had fewer than 5 animals within a year throughout the analysed period; the number of animals tested per month within a year was less than 20. Altogether, 41 records were removed. CGs were defined as herd-year-season in the last year of analysis.

Descriptive statistics were obtained using SAS software [SAS inc. 2004]. Data were analysed using the following linear mixed model:

$$y = Xb + Z_c c + Z_a a + e \qquad (1)$$

where:

y – vector of observations for BF, b is a vector of unknown fixed effects (sex, herd-year-season, where season was defined as year-month of testing, genotype, and weight at the end of test);

c and a – the vectors of unknown the random effects of common litter environment and breeding values, respectively;

e – vector of residuals, while X, Z_c, and Z_a are design matrices linking phenotype records with corresponding effects. Model assumptions were: $E(y) = Xb$, $c \sim N(0, I\sigma^2_c)$, $a \sim N(0, A\sigma^2_a)$, $e \sim N(0, I\sigma^2_e)$, where A is the numerator relationship matrix.

Gene flow

GF was used as a measure of connectedness to show the amount and direction of genetic links between herds under the breeding program, as proposed by Kennedy and Trus [1993]. A measure of connectedness can be provided by the X'ZTQ multiplication, where Q identifies basal animals with respect to their herd of origin, and T is a lower triangular matrix that traces the flow of genes from one generation to the next, so that the numerator-relationship matrix A=T'WT, and W is a diagonal matrix of Mendelian sampling variances. The average GF in the ten-year period between the CGs was calculated. GF between CGs was calculated for L, LW, and for the data set that contained both breeds and their crossbreds.

Connectedness rating

In addition, the level of connectedness between CGs was described using the connectedness rating (CR) measure [Mathur *et al.* 1998], which is defined as the correlation (2) between the estimates for CG i $(\hat{h}_i)$ and CG j $(\hat{h}_j)$:

$$CR_{i,j} = \frac{Cov(\hat{h}_i, \hat{h}_j)}{\sqrt{Var(\hat{h}_i)Var(\hat{h}_j)}} \qquad (2)$$

where: variances and covariances were obtained by direct solution of the left-hand side of mixed model equations (MME) (3). The MME for the model (1) are as follows:

$$\begin{bmatrix} \mathbf{X^T X} & \mathbf{X^T Z_c} & \mathbf{X^T Z_a} \\ \mathbf{Z_c^T X} & \mathbf{Z_c^T Z_c} + \mathbf{I}\sigma_e^2\sigma_c^{-2} & \mathbf{Z_c^T Z_a} \\ \mathbf{Z_a^T X} & \mathbf{Z_a^T Z_c} & \mathbf{Z_a^T Z_a} + \mathbf{A}^{-1}\sigma_e^2\sigma_a^{-2} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{c} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X^T y} \\ \mathbf{Z_c' y} \\ \mathbf{Z_a' y} \end{bmatrix} \qquad (3)$$

The average CR between CGs in the data set was calculated in the last year Also, CR was calculated within CGs. The calculation was performed for different combinations of data in the sets. In the first case, CR between CGs was calculated using purebred data separately for L and LW. Further, CR between CGs was calculated using a joint data set for L and LW. Finally, CR was calculated using data for both pure breeds and crossbreeds.

Reliability of estimated breeding values (r^2) was calculated as:

$$r^2 = 1 - (PEV / \sigma_a^2) \qquad (4)$$

where PEV is the prediction error variance and σ_a^2 is additive genetic variance in the population [e.g. Mrode, 2005]. PEV was obtained by the direct solution of the left-hand side of MME (3).

In order to estimate breeding values and their reliabilities, components of variance were estimated. The same values for variance components were used to estimate breeding values and their reliability in order to ensure an adequate comparison of breeding values between CGs. For this purpose, the phenotypic variance of the BF was divided into the following components: common litter environmental effect, additive genetic effect and residual. Reliability was calculated for all breeding values and average reliability was calculated for every single management unit and breed both separately and jointly. Also, average reliability was calculated for breeding values in a joint analysis of all farms and breeds. Finally, the average reliability that included data records was calculated for crossbreds within and across the herds.

The residual maximum likelihood implemented in the VCE 6 software [Groeneveld *et al.* 2008] was applied to estimate variance components for all large farms and the group of family farms, as well as variance components for the whole analysed population. The PEST program [Groeneveld *et al.* 1990] was used to predict breeding values (BV).

Results and discussion

Breed structure and descriptive statistics of the analysed data sets are shown in Table 1. The breed with the highest proportion in the analysed population was L (33%), while crossbreds between L and LW had the lowest proportion (16%). The average GF in the period of 2002-2012 for L and LW is shown in Figure 1 and Figure 2. The genetic contribution of animals from one particular herd to another is expressed as a proportion of genes in the CG originating from other CGs involved in genetic evaluation. For both breeds, the proportion of genes in a particular CG originating from other CGs was low. The highest proportion of genes originating from other CGs

Table 1. Number of animals and descriptive statistics by genotype and large scale (A-C) and a group of family farms (F)

Farm	N	Mean (BF, mm)	SD	L	LW	♀ LW x L ♂	♀ L x LW ♂
A	15,204	11.86	2.32	3,327	3,966	4,195	3,716
B	4,610	11.22	1.88	1,579	675	2,356	/
C	2,761	11.55	2.42	2,342	82	337	/
F	670	10.28	2.20	422	120	128	/
Total	23,245			7,670	4,843	7,016	3,716

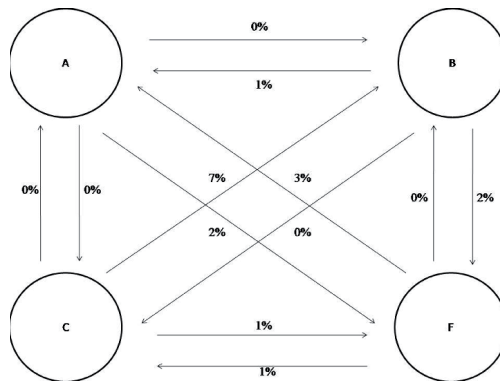L – Landrace; LW-Large White.

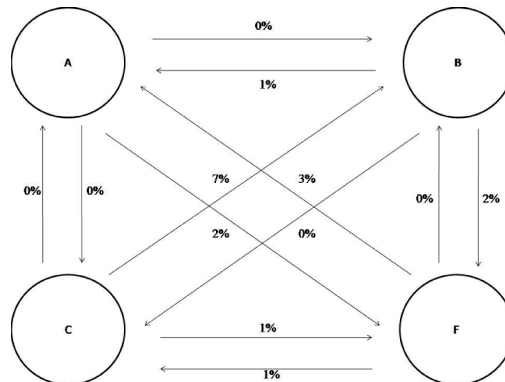Fig. 1. Gene flow between contemporary groups for Landrace.

Fig. 2. Gene flow between contemporary groups for Large White.

was found in the L, where 7% of genes on farm B originated from farm C. A greater proportion of genes exchanged between farm A and the group of family farms was found in LW, where GF was determined in both directions, with proportions of 6% and 4%, respectively. In several cases there was no GF between the herds in both breeds. The low proportion of genes from other CGs participating in the across-herd
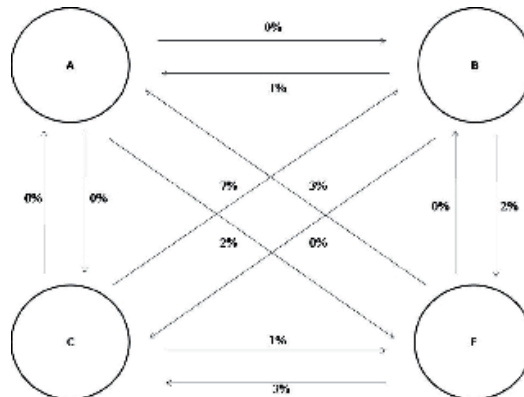
Fig. 3. Gene flow between contemporary groups for joint Landrace, Large White and their crossbreeds.

Table 2. Connectedness rating (CR) between contemporary groups for Landrace

Farm	A	B	C	F
A	0.019	0.002	0.001	0.001
B	/	0.020	0.003	0.002
C	/	/	0.020	0.003
F	/	/	/	0.029

A-C – large scale farms; F – group of family farms.

comparison implied a low level of connectedness between them. A similar situation was observed when GF was analysed in the data set that contained records of both purebreds and crossbreds (Fig. 3). The proportion of genes from farm C found on farm B was 7%. Proportions of genes originating from other farms on farm B were within the range of 0-3%. Thus, crossbreeding was performed within herds and there was no exchange of purebred animals between herds in order to produce crossbreeds. The levels of connectedness between CGs for L are shown in Table 2. Values of CR between CGs were low and did not approach 3%, which is the minimal value of CR needed for across-herd genetic evaluation, as suggested by Mathur *et al.* [2002]. The highest CR value was very close to the desired 3% in the group of family farms, suggesting that the across-herd comparison within this group could be reliable with a small additional increase in CR. Similarly, low values for CR could be observed in the LW breed (Tab. 3), except for CR between farms A and the group of family farms, where a small additional increase in CR would enable a reliable across-herd evaluation. The CR values for LW between other farms were low and did not reach 3% and thus, a comparison of genetic evaluation across herds was not recommended for both pure breeds. The reasons for such a low GF and CR between CGs might be explained by the fact that large farms included in the breeding program have highly intensified, but usually isolated systems without a direct exchange of genetic material

Table 3. Connectedness rating (CR) between contemporary groups for Large White

Farm	A	B	C	F
A	0.056	0.010	/	0.025
B	/	0.020	/	0.001
C	/		/	/
F	/	/	/	0.029

A-C – large scale farms; F – group of family farms.

with other farms. Moreover, the low level of connectedness between family farms and large farms for L are a consequence of the relatively low proportion of the L breed on family farms. Large farms buy semen from AI centres, but they also performed their own selection during the analysed period, and this genetic progress was not transferred/shared between them.

The values of CR within CGs were higher when compared with CGs from different management units. The observed difference is due to selection and matings within herds. However, the level of connectedness required for the across-herd comparison was achieved only within the group of family farms for both breeds and in farm A for the LW.

Values of CR between management units when data set for both pure breeds were used are shown in Table 4. Values of CR between farm A and the group of family farms and farms A and B were higher than CRs between other farms, due to the high CR for LW. This result indicates that additional efforts may result in an increase of the

Table 4. Connectedness rating (CR) between contemporary groups for Landrace and Large White

Farm	A	B	C	F
A	0.099	0.012	0.008	0.015
B	/	0.018	0.004	0.002
C	/	/	0.019	0.002
F	/	/	/	0.015

A-C – large scale farms; F – group of family farms.

Table 5. CR between contemporary groups with purebred and crossbred data included

Farm	A	B	C	F
A	0.066	0.020	0.016	0.015
B	/	0.024	0.009	0.004
C	/	/	0.020	0.004
F	/	/	/	0.130

A-C – large scale farms; F – group of family farms.

connectedness, and consequently, to enable across-herd genetic evaluation for a joint analysis of both breeds. No substantial change in CR between herds was noticed when crossbred data were added to the analysis (Tab. 5). A slight increase observed was not sufficient to enable a reliable across-herd comparison.

The highest CR was noticed between farms A and B, and farm A and the group of family farms. A substantial increase when crossbred data were included in the data set was noticed in the group of family farms. The probable reason for this increase is due to fact that exchange of live animals between herds within this group was more intense than between the large farms and crossbred production by exchanging genetic materials among herds.

Although the level of connectedness between CGs was low, weak genetic associations between them could still be found. The existing connectedness is probably a consequence of using same boars or semen from AI centres in the past. Links between family farms and farm A were probably formed through the direct transfer of live animals from large farms to family farms. Also, it is possible that the direct transfer of genetically superior live animals from family farms to large farms existed at some points of the analysed period. Although genetic links are weak, their presence in the population can be used as a starting point for increasing the level of connectedness. Mathur *et al.* [2002] recommended that for BF average connectedness rating between a certain management unit and all other units should be 3% to ensure a reasonably reliable comparison of estimated breeding values between management units. According to this criterion, the results of genetic evaluation within the breeding program should not be compared across all management units for the analysed breeds, because direct and average connectedness ratings were never higher than 3% for any CG. An adequate tool for increasing the level of connectedness between CGs is to use common boars in herds and increase the proportion of litters of common boars within management units, as proposed by Sun *et al.* [2009]. Additional efforts to achieve an adequate organisation of AI within breeding programs are also necessary to ensure a simultaneous improvement of connections between CGs. In order to increase the level of connectedness and enable adequate comparisons of estimated breeding values including data for both breeds and their crossbreds, farms should allow exchange of purebred animals to produce crossbreds.

Although the low level of connectedness between management units affects the accuracy of comparisons between estimated breeding values, it may also have an effect on the accuracy or reliability of every single breeding value. Thus, the reliability of estimated breeding values from across-herd genetic evaluation depends on the amount of information included in the evaluation, as well as the level of connectedness between management units in the across-herd evaluation. In the case of well connected herds, the reliability of estimated breeding values is expected to be higher when joint evaluation is done, due to the increase in the number of phenotype records included in the evaluation and contribution of information from one herd to other herds. However, if herds are poorly connected, no meaningful increase in the

Table 6. Average reliability of prediction within contemporary groups using purebred data

Farm	Data set					
	DS1		DS2		DS3	
	L	LW	L	LW	L	LW
A	0.32	0.33	0.34	0.34	0.37	0.36
B	0.24	0.22	0.25	0.23	0.27	0.28
C	0.28	0.03	0.31	0.05	0.29	0.07
F	0.08	0.05	0.09	0.07	0.09	0.06

DS1 – reliability of genetic evaluations within breeds and management units; DS2 – reliability of genetic evaluations for LA and LW within management units; DS3– reliability of genetic evaluations for purebreds and crossbreds within management units.

Table 7. Average reliability of prediction in contemporary groups using purebred and crossbred data

Farm	Data set					
	DS4		DS5		DS6	
	L	LW	L	LW	L	LW
A	0.38	0.38	0.39	0.39	0.43	0.42
B	0.32	0.27	0.34	0.33	0.38	0.38
C	0.34	0.09	0.35	0.26	0.35	0.31
F	0.24	0.13	0.25	0.19	0.27	0.21

DS4 – reliability of across-herd genetic evaluations within breed; DS5 – reliability of across-herd genetic evaluations for both purebreds; DS6 – reliability of across-herd genetic evaluations for purebreds and crossbreds.

reliability of across-herd genetic evaluation is expected, as information from one CG does not contribute to other CGs, due to the lack of genetic links. Despite the low level of connectedness, a slight increase in the reliability of estimated breeding values was noticed within the management units with a large number of data records when an additional amount of data was included in the analysis (Tab. 6 and 7). A remarkable increase in reliability was observed in CGs with a smaller number of observations, especially in the group of family farms. The difference in average reliabilities in the group of family farms ranged from 0.08 to 0.27 for L and from 0.03 to 0.31 for LW on farm C. The increase in the reliability of estimated breeding values in all CGs was probably due to the greater amount of information when a joint evaluation was performed. The highest increase was observed for the group of family farms, which had the smallest number of records, but were well connected with farm A, which had the largest number of records. Increased reliability is consistent with the results obtained in a study by Ibáńez-Escrische *et al.* [2011], where a similar pattern was also observed when crossbred data were added into the genetic evaluation of the BF in pigs. Considering the low connectedness level between all management units, there is still space for an additional increase of individual reliability in the across-herd

comparison. Although the increase in average reliability within management units, i.e. the average for individual breeding values was observed when additional crossbred data were introduced, low levels of connectedness suggested that the comparison of estimated breeding values between herds might be unreliable.

This study showed that the level of connectedness between CGs obtained in all data combinations was low and not sufficient for a reliable across-herd comparison. However, the reliabilities of estimated breeding values were slightly increased by including crossbred data into the analysis due to an increased amount of data. The accuracy of the across-herd comparisons throughout the breeding program may thus be additionally improved by increasing the level of connectedness between management units. Proper organization of AI and exchange of purebred animals in order to produce crossbreeds are recommended in order to achieve higher levels of connectedness between management units in pigs.

REFERENCES

1. GROENEVELD E., KOVAC M., MIELENZ, N., 2008 – VCE 6**,** User's guide 218 and 219 manual, Version 6.0.
2. GROENEVELD E., KOVAC M.,WANG T., 1990 – PEST, a general purpose BLUP package for multivariate prediction and estimation. 4th World Congress on Genetics Applied to Livestock Production, 23-27 July, 1990, Edinburgh, 13, 488-491.
3. HENDERSON C.R., 1949 – Estimates of changes in herd environment. *Journal of Dairy Science*, 32, 706 (abstract).
4. IBÁŃEZ-ESCRISCHE, N., REIXACH, J., LLEONART, N., NOGUERA, J.L., 2011 – Genetic evaluation combining purebred and crossbred data in pig breeding scheme. *Journal of Animal Science* 89**,** 3881-3889.
5. KENNEDY, B.W., TRUS, D., 1993 – Considerations on genetic connectedness between management units under an animal model. *Journal of Animal Science* 71, 2341-2352.
6. LEWIS, R.M., CRUMP, R.E., KUEHN, L.A., SIMM, G., THOMPSON, R., 2005 – Assessing connectedness in across-flock genetic evaluations. Proceedings of the British Society of Animal Science, 22-24 March, Scarborough, 121-122.
7. MATHUR, P.K., SULLIVAN, B., CHESNAIS, J., 1998 – A new method for assessing connectedness between herds. Proceedings of the National Swine Improvement Federation Conference and Annual Meeting, 5 December 1998, East Lansing, Michigan 23, 33-39.
8. MATHUR, P.K., SULLIVAN, B.P., CHESNAIS, J.P., 2002 – Measuring connectedness: concept and application to a large industry breeding program. Proceeding of the 7th World Congress of Genetics Applied to Livestock Production, August 19-23, Montpellier, France, 32, 545-548.
9. MRODE, R.A., 2005 – Linear models for the prediction of animal breeding values. CABI Publishing, Wallingford, UK pp. 51.
10. SAS INST. INC., 2004 – 215 The SAS System for Windows. Release 9.1.3. Cary, NC.
11. SMITH, C., BANOS, G., 1991 – Selection within and across populations in livestock improvement. *Journal of Animal Science* 69, 2387-2394.
12. SUN, C.Y., WANG C.K., WANG Y.C., ZHANG Y., ZHANG Q., 2009 – Evaluation of connectedness between herds for three pig breeds in China. *Animal* 3, 482-485.
13. VESELÁ, Z, PŘIBYL, J, VOSTRÝ L, ŠTOLC, L. 2007 – Stochastic simulation of the influence of insemination on the estimation of breeding value and its reliability. *Czech Journal of Animal Science* 52**,** 236-248.