

## GENETICS AND BREEDING

### Calculation and Use of Inbreeding Coefficients for Genetic Evaluation of United States Dairy Cattle

G. R. WIGGANS and P. M. VanRADEN  
Animal Improvement Programs Laboratory  
Agricultural Research Service, USDA  
Beltsville, MD 20705-2350

J. ZUURBIER<sup>1</sup>  
Department of Animal Sciences  
University of Maryland  
College Park 20742

#### ABSTRACT

Inbreeding coefficients are calculated routinely for all animals included in national genetic evaluations for yield traits. The base population for inbreeding is defined as animals born during 1960. Animals with unknown parents are assumed to have inbreeding coefficients that are equal to the mean of coefficients for animals with known parents born during the same year. Mean inbreeding coefficients reached .03 to .04 for recent years, and coefficients for some animals exceeded .50. The annual increase in level of inbreeding was highest for Milking Shorthorns, but the rate of change of that increase was greatest for Holsteins. Accounting for inbreeding in calculation of the inverse of the relationship matrix had only a small effect on evaluations. For Jerseys, the maximum change in breeding value was 73 kg of milk for cows and 40 kg of milk for bulls with  $\geq 10$  daughters. Estimates of inbreeding depression were similar across breeds for production traits and were -29.6 kg of milk, -1.08 kg of fat, and -.97 kg of protein per 1% of inbreeding for Holsteins. In January 1994, the USDA began considering the percentage of inbreeding when calculating inverses of relationship matrices, the largest matrix representing over 20 million Holsteins; this inbreed-

ing percentage was released to the dairy industry for bulls.

(Key words: inbreeding, relationship matrix, genetic evaluation)

Abbreviation key: ID = identification.

#### INTRODUCTION

Although inbreeding occurs in all populations to some degree, the effects of inbreeding generally have been ignored for genetic evaluation of US dairy cattle, primarily because of computational difficulties related to large population size. Inbreeding reduces phenotypic performance (inbreeding depression) and the genetic variation due to Mendelian sampling among progeny of a particular set of parents. Measures of inbreeding (inbreeding coefficients) represent the homozygosity expected in an animal's genome because of the relatedness of its parents.

The effect of inbreeding on mean phenotypic performance can be accounted for in genetic evaluations by regression on the percentage of inbreeding. An inbreeding effect can be routinely included in the model, or data can be preadjusted by a regression coefficient that is estimated separately.

Adjustment of the inverse of the relationship matrix to include inbreeding coefficients can account for the effect of inbreeding on genetic variance. Quaas (8) proposed a method for constructing this matrix for large populations.

Inbreeding also affects estimates of the accuracy (reliability) of genetic evaluations. Boichard and Lee (1) reported increased overestimation of accuracy with increases in in-

Received September 6, 1994.

Accepted March 13, 1995.

<sup>1</sup>Current address: Department of Animal Breeding, Agricultural University, Wageningen, The Netherlands.

breeding when applying the method of Meyer (6) to approximate accuracy of evaluation under an animal model. This overestimation resulted from accounting for the several contributions that an inbred animal has from the same ancestor as though they were contributions from independent ancestors.

Many animals with information that is included in genetic evaluations have incomplete pedigrees reported. Estimates of inbreeding for these animals can be improved by considering unknown parents to be related instead of unrelated (12). Genetic evaluations rank animals and are expressed relative to a base population, and a base population also is necessary for inbreeding coefficients (3, 12). The base is arbitrary. The inbreeding coefficient measures the increase in homozygosity that is due to common ancestors born since the base year. If the base is set far enough in the past so that recent common ancestors are considered, evaluations of current animals are not sensitive to which base year was chosen.

Advances in computer technology and computational methods have enabled the rapid calculation of inbreeding coefficients for large populations (4, 5, 7, 11, 12). An algorithm of VanRaden (12) was used to calculate inbreeding coefficients for 9 million registered Holsteins. Further work was required to integrate this method into routine national evaluations and to estimate inbreeding coefficients for nonregistered animals by incorporation of unknown-parent groups.

The purpose of this study was 1) to describe how inbreeding coefficients are calculated for US national evaluations, 2) to document how the US genetic evaluation system for yield traits was modified to account for the effect of inbreeding on genetic variance, 3) to determine the effect on bull evaluations of accounting for inbreeding in the inverse of the relationship matrix, and 4) to estimate inbreeding depression.

**MATERIALS AND METHODS**

Pedigree data for calculation of inbreeding coefficients were obtained from herdbook information provided by breed associations for registered Ayrshires, Brown Swiss, Guernseys, Holsteins, Jerseys, and Milking Shorthorns. For nonregistered cows, pedigree data were obtained from the lactation records reported by

dairy records processing centers. The percentage of known ancestors for Holstein cows born during 1990 is shown in Figure 1. All cow sires were known because sire identification was required for inclusion of cows in the data. For each of the five generations included, a greater percentage of sires than dams was known. The smallest percentage of known ancestors was 49 for maternal granddams of maternal granddams.

Inbreeding calculations were separated into two computer programs to minimize memory requirements. The first program used the hashing procedure of Wiggans et al. (14) to recode the permanent identification (ID) (registration or eartag number) of bulls and cows to a set of sequential numbers. This recoding allowed pedigree data for an entire breed to be stored in memory. Incorrect or impossible parent ID (e.g., identical sire and animal ID) was designated as unknown. Only parents of animals born during 1960 or later were included because 1960 was defined as the base year for inbreeding (12). Complete pedigree informa-

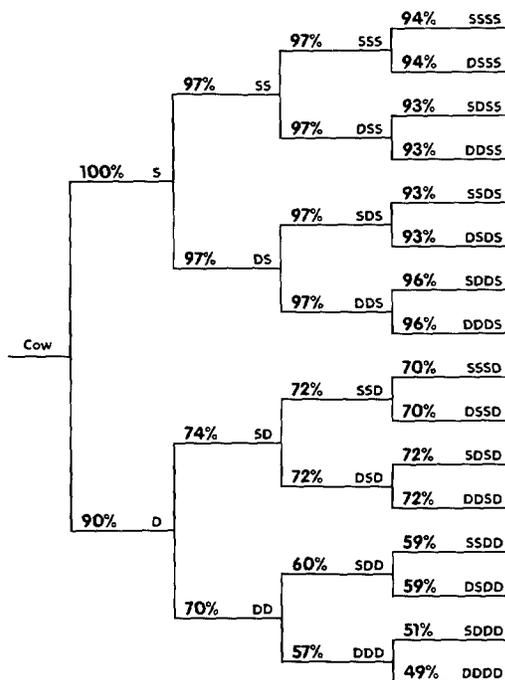


Figure 1. Percentage of known ancestors for Holstein cows born during 1990 (S = sire and D = dam; e.g., DS is the dam of the sire).

tion for animals born before 1960 was not always available in an electronic format. Animals born before 1960 were assumed to be unrelated, and their pedigree data were excluded. Records were processed in the sequence of animal ID with cows processed before bulls. As each animal was processed, parent ID was recoded (hashed) before animal ID.

The second computer program stored the hashed pedigree data, birth year, and gender in memory. For parents with unknown birth years, birth year was estimated as 3 yr prior to the earliest birth year of progeny. Because inbreeding coefficients based on birth year are required for unknown parents at the beginning of the relationship matrix for each animal, inbreeding coefficients for all animals from the previous semiannual evaluation were averaged for each birth year, multiplied by 2, and used as starting values in the matrix. This procedure requires iteration for the annual mean to stabilize. However, once the mean has stabilized, no iteration is required for subsequent inbreeding calculations. The partition for the inbreeding coefficients of unknown parents was of order 36 (the current year, 1994, minus 1958). Each diagonal element was propagated to the other elements in its row and column associated with earlier years. This procedure caused the relationship between unknown parents born during different years to be that of animals born during the birth year of the most recent unknown parent, which differs from the procedure of VanRaden (12) in which unknown parents are assumed to be related to all other parents by twice the mean inbreeding level of the period.

A relationship matrix was created for each animal following the method of VanRaden (12). Animals were processed in descending sequential number order so that progeny would be processed before parents. An animal's inbreeding coefficient was not recalculated if it already had been calculated as an ancestor. Information for all ancestors of an animal was collected by formation of a vector that was long enough for the largest family. The vector was filled by addition of information from each ancestor's parents at the end of the vector as each ancestor was processed. Processing ended because ancestors with unknown parents did not add information to the vector. The

ancestry vector was filled, starting with the highest number. For duplicate numbers, the entry lowest in the vector was retained while elements of the vector were copied to its lower end; the lowest positions were occupied by the unknown-parent groups. Another vector also was created to store the location of parents of a particular animal in the ancestry vector. For unknown ancestors, the unknown-parent group that was appropriate for that birth year was selected. Animals from the most recent 3 yr of birth were combined into a single group because those few animals were primarily young, registered, nonlactating, and unrepresentative of the milking population.

The relationship matrix was filled starting with the first animal above the birth-year groups for unknown parents. The inbreeding coefficient was one-half the relationship between the parents, and the relationship of the animal with itself was 1 plus the inbreeding coefficient. Mean inbreeding coefficients were calculated for each birth year. The first time that the inbreeding coefficient was calculated for an animal, that coefficient contributed to the year total.

After inbreeding coefficients were calculated for all animals, the birth-year means were written to a file. Files of pedigree and hash numbers were read, and the pedigree files were rewritten with inbreeding added.

The inbreeding coefficients were used in the iteration program to construct the inverse of the relationship matrix. To determine an animal's contribution to its diagonal, the ratio of error variance to additive genetic variance ( $\alpha$ ) was adjusted to reflect the reduced variance of Mendelian sampling that results from inbreeding and to reflect whether or not a parent was known (12):  $\alpha[4/(2 - F_s - F_d)]$  if both parents were known,  $\alpha[4/(3 - F_s - 2F_d)]$  if only the sire was known,  $\alpha[4/(3 - 2F_s - F_d)]$  if only the dam was known, and  $\alpha[2/(2 - F_s - F_d)]$  if neither parent was known, where  $F_s$  and  $F_d$  are the inbreeding coefficients of the sire and dam, respectively. For parents, .25 times this contribution was added to the diagonals for sire and dam. Off-diagonals between animal and parents received  $-.5$  times this contribution; off-diagonals between parents received .25 times this contribution (12). Two vectors were maintained: the diagonals and the Mendelian sampling. The second vector was required for sub-

TABLE 1. Numbers of animals born during 1960 or later and their parents, maximum numbers of ancestors per pedigree, and processing times required to calculate inbreeding coefficients for major US breeds of dairy cattle.

Breed	Animals and parents	Maximum ancestors per pedigree	Processing time <sup>1</sup>
	(no.)		(min)
Ayrshire	381,702	272	3.2
Brown Swiss	406,171	289	4.5
Guernsey	1,484,269	310	9.8
Holstein	20,869,458	380	277.9
Jersey	1,944,089	331	20.4
Milking Shorthorn	74,497	275	1.1

<sup>1</sup>Processing time on an IBM (Armonk, NY) RISC System 6000 POWERserver computer (model 560).

tracting the off-diagonals from the right-hand sides while iterating on the data.

The effect of properly accounting for inbreeding when constructing the inverse of the relationship matrix was investigated for the Jersey breed. Evaluations resulting from using correct inbreeding coefficients and those with inbreeding assumed to be 0 were compared for cows and AI bulls.

To estimate inbreeding depression, a linear regression on inbreeding was added to the computer program that calculates solutions through iteration (14, 15, 17). Addition of this factor to the model for analysis ensured that the estimates of inbreeding depression would be consistent with other adjustments to the data, such as those for heterogeneous variance (16), and that genetic merit would be estimated simultaneously. A linear effect was chosen following preliminary investigation of quadratic and categorical effects. With the more complete models, high levels of inbreeding sometimes showed less inbreeding depression than lower levels. The reliability of these estimates was low because of the small number of animals that were highly inbred.

## RESULTS

The computing time required to calculate inbreeding coefficients for each breed is in Table 1. The number of ancestors per pedigree (Table 1) is an indicator of computing time, which increases with family size and was greatest for Holsteins. For the 21 million Holsteins, the two programs used to calculate inbreeding coefficients required <5 h. Because

the computing time was short, recalculation of all inbreeding coefficients was preferred to calculation of coefficients only for new animals or for animals with corrected pedigrees. An updating system would have required identification of all descendants of animals with corrected pedigrees.

Mean inbreeding coefficients are in Figure 2 for birth years  $\geq 1960$ . All breeds started with inbreeding coefficients of 0 in 1960 because of the base definition and accumulated inbreeding since then. Mean inbreeding coefficients for animals born during 1990 (Table 2) ranged from .026 for Holsteins to .047 for Ayrshires. Based on a quadratic curve fit to the inbreeding means, annual increase in inbreeding (first



Figure 2. Mean inbreeding coefficients for Ayrshires (A), Brown Swiss (B), Guernseys (G), Holstein (H), Jerseys (J), and Milking Shorthorns (M).

TABLE 2. Mean inbreeding coefficients for animals born during 1990 from a quadratic prediction equation with first and second derivatives for major US breeds of dairy cattle.

Breed	Animals (no.)	$\bar{X}$	First derivative	Second derivative ( $\times 10^{-5}$ )
Ayrshire	20,605	.047	.00172	.099
Brown Swiss	27,729	.030	.00172	4.920
Guernsey	104,951	.035	.00220	7.344
Holstein	1,136,123	.026	.00203	8.334
Jersey	112,562	.033	.00210	7.307
Milking Shorthorn	7948	.041	.00245	7.518

derivative) was highest for Milking Shorthorns, but rate of change of that increase (second derivative) was greatest for Holsteins. Short et al. (10) reported mean inbreeding coefficients of .032 for cows and .035 for bulls for registered US Holsteins born during 1990.

Because the population base for inbreeding was 1960, the number of ancestors was affected by level of inbreeding and by generation interval. Table 3 shows the distribution of inbreeding coefficients for Holstein cows born during 1990. The highest inbreeding coefficient was .384; the mode was .019. Among all breeds, the highest inbreeding coefficient was .56 for a Guernsey bull. Inclusion of unknown-parent groups increases the number of animals with low inbreeding.

Table 4 shows the frequency of differences between breeding values for milk calculated with or without inbreeding coefficients that were included in the inverse of the relationship matrix for Jerseys. For most Jersey cows, inclusion of inbreeding coefficients was of no consequence in estimation of breeding value. The difference (breeding value with inbreeding minus breeding value without inbreeding) of greatest magnitude was -73 kg for cows and -40 kg for bulls with  $\geq 10$  daughters. Animals with breeding values that changed when correct inbreeding coefficients were used had one or both parents with high inbreeding coefficients. Casanova et al. (2) also found that use of correct inbreeding coefficients in the relationship matrix of Swiss Braunvieh cattle had little effect on predicted breeding values for milk yield and generally did not change the

TABLE 3. Distribution of inbreeding coefficients for Holstein cows born during 1990.

Inbreeding coefficient	Cows	
	(no.)	(%)
0	14,705	1.66
.01	218,040	24.56
.02	286,006	32.21
.03	186,511	21.01
.04	76,952	8.67
.05	41,211	4.64
.06	21,194	2.39
.07	16,820	1.89
.08	10,797	1.22
.09	5563	.63
.10	2854	.32
.11	1086	.12
.12	408	.05
.13	1553	.17
.14	1214	.14
.15	692	.08
.16	415	.05
.17	201	.02
.18	71	.01
.19	53	.01
.20	23	<.01
.21	3	<.01
.22	5	<.01
.23	3	<.01
.24	2	<.01
.25	234	.03
.26	840	.09
.27	222	.03
.28	77	.01
.29	28	<.01
.30	8	<.01
.31	1	<.01
.32	4	<.01
.33	0	0
.34	0	0
.35	0	0
.36	0	0
.37	0	0
.38	13	<.01

ranking of animals. The current method (13) for estimation of reliability is expected to overestimate the reliability of inbred animals because parent information is assumed to be independent. Changes in the reliability of evaluations from use of correct inbreeding coefficients were expected to be small and were not investigated.

Estimates of inbreeding depression are in Table 5 for milk, fat, and protein yields. Estimates were similar across breeds for each production trait. For Holsteins, estimates of -29.6 kg of milk, -1.08 kg of fat, and -.97 kg of protein per 1% of inbreeding were slightly larger than the -22.6, -.78, and -.85 kg reported by Short et al. (10) for registered Holsteins. For Jerseys, estimates of -21.3 kg of milk and -1.03 kg of fat per 1% inbreeding

were much larger than those reported for Canadian Jerseys (-9.84 kg and -.55 kg) by Miglior et al. (7). Accounting for inbreeding depression is expected to increase estimates of genetic trend by the amount of inbreeding depression times the trend in inbreeding. Trend then would measure the annual genetic improvement for animals that were not inbred.

CONCLUSIONS

Inbreeding coefficients can be calculated routinely for the US dairy population. In January 1994, the USDA began calculating the percentage of inbreeding for all animals, and this percentage was released to the dairy industry for bulls (9). Inclusion of inbreeding coefficients in the inverse of the relationship matrix resulted in a more accurate representation of the relationships among animals. This adjustment of genetic variance affects all traits (yield, productive life, and somatic cell score) evaluated by the USDA with an animal model. However, accounting for inbreeding had relatively little impact on yield evaluations. Its impact may increase as inbreeding increases.

Accounting for inbreeding depression has not been implemented by the USDA in national evaluations, although inbreeding depression is expected to have a greater influence on evaluations than does use of the correct inbreeding coefficients in calculating the inverse of the relationship matrix. Accuracy of selection would improve if evaluations were adjusted to include the inbreeding depression expected with each mate or population of mates. Accounting for inbreeding depression also would prevent a young bull from being penalized during his initial evaluation for matings to relatives. However, bulls with highly inbred

TABLE 4. Frequency of differences<sup>1</sup> between breeding values for milk calculated with correct inbreeding coefficient or with no inbreeding coefficient included in the inverse of the relationship matrix for Jersey cows and bulls with ≥10 daughters.

Difference (kg)	Cows	Bulls
-73 to -70	1	0
-69 to -65	2	0
-64 to -60	0	0
-59 to -55	2	0
-54 to -50	3	0
-49 to -45	4	0
-44 to -40	14	1
-39 to -35	28	1
-34 to -30	44	0
-29 to -25	128	0
-24 to -20	296	0
-19 to -15	926	3
-14 to -10	3536	18
-9 to -5	19,495	89
-4 to 0	573,614	5090
1 to 5	435,562	4111
6 to 10	14,684	63
11 to 15	3050	8
16 to 20	849	5
21 to 25	280	0
26 to 30	137	1
31 to 35	42	0
36 to 40	26	0
41 to 45	11	0
46 to 50	3	0
51 to 55	0	0
56 to 60	1	0
61 to 65	0	0
66 to 68	1	0

<sup>1</sup>Breeding value with correct inbreeding coefficient minus breeding value with no inbreeding assumed.

TABLE 5. Estimates of inbreeding depression for production traits of major US breeds of dairy cattle.

Breed	Milk	Fat	Protein
	(kg/1% inbreeding)		
Ayrshire	-30.2	-1.16	-1.20
Brown Swiss	-24.6	-1.08	-.99
Guernsey	-19.6	-.89	-.77
Holstein	-29.6	-1.08	-.97
Jersey	-21.3	-1.03	-.88
Milking Shorthorn	-22.0	-.68	-1.01

daughters may be highly related to the population of potential mates. Therefore, high inbreeding and inbreeding depression would be expected in future progeny. To avoid overestimation of a bull's genetic merit when he is bred to related mates, a useful statistic could be the bull's mean relationship with expected mates. Accounting for inbreeding depression in national genetic evaluations probably will not be implemented until such a statistic is calculated routinely and until mating programs consider inbreeding depression.

The adoption of an animal model for genetic evaluations could have contributed to the increase in inbreeding in recent years. Because of full consideration of relationships, related animals tend to rank together. Continued increase in inbreeding will cause greater inbreeding depression. Selection and mating programs that directly account for inbreeding automatically will choose and pair less related parents.

#### ACKNOWLEDGMENTS

Cooperation of the US dairy industry in supplying yield and pedigree data through the National Cooperative Dairy Herd Improvement Program is appreciated. Manuscript improvements were suggested by S. M. Hubbard, Animal Improvement Programs Laboratory, Beltsville, MD, and the reviewers of the *Journal of Dairy Science*.

#### REFERENCES

- 1 Boichard, D., and A. J. Lee. 1992. Approximate accuracy of genetic evaluation under a single trait animal model. *J. Dairy Sci.* 75:868.
- 2 Casanova, L., C. Hagger, N. Kuenzi, and M. Schneeberger. 1992. Inbreeding in Swiss Braunvieh and its influence on breeding values predicted from a repeatability animal model. *J. Dairy Sci.* 75:1119.
- 3 Falconer, D. S. 1989. *Introduction to Quantitative Genetics*. 3rd ed. Longman, Inc., New York, NY.
- 4 Golden, B. L., J. S. Brinks, and R. M. Bourdon. 1990. A performance programmed method for computing inbreeding coefficients from large data sets for use in mixed-model analyses. *J. Anim. Sci.* 69:3564.
- 5 Meuwissen, T.H.E., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. *Genet. Sel. Evol.* 24:305.
- 6 Meyer, K. 1989. Approximate accuracy of genetic evaluation under an animal model. *Livest. Prod. Sci.* 21:87.
- 7 Miglior, F., B. Szkotnicki, and E. B. Burnside. 1992. Analysis of levels of inbreeding and inbreeding depression in Jersey cattle. *J. Dairy Sci.* 75:1112.
- 8 Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* 32:949.
- 9 Schutz, M. M., P. M. VanRaden, and G. R. Wiggans. 1994. Recent improvements in the USDA-DHIA genetic evaluation system for dairy cattle. Page 45 in *New Uses for DHIA Data*. Natl. DHIA, Columbus, OH.
- 10 Short, T. H., T. J. Lawlor, and R. W. Everett. 1992. Inbreeding in U.S. Holsteins and its effect on yield and type traits. *J. Dairy Sci.* 75(Suppl. 1):154.(Abstr.)
- 11 Tier, B. 1990. Computing inbreeding coefficients quickly. *Genet. Sel. Evol.* 22:419.
- 12 VanRaden, P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of large populations. *J. Dairy Sci.* 75:3136.
- 13 VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. *J. Dairy Sci.* 74:2737.
- 14 Wiggans, G. R., I. Misztal, and L. D. Van Vleck. 1988. Implementation of an animal model for genetic evaluation of dairy cattle in the United States. *J. Dairy Sci.* 71(Suppl. 2):54.
- 15 Wiggans, G. R., and P. M. VanRaden. 1990. Including information from records in later herds in animal model evaluations. *J. Dairy Sci.* 73:3336.
- 16 Wiggans, G. R., and P. M. VanRaden. 1991. Method and effect of adjustment for heterogeneous variance. *J. Dairy Sci.* 74:4350.
- 17 Wiggans, G. R., and P. M. VanRaden. 1993. Flow of information for genetic evaluation of yield traits. Page 19 in *Proc. Symp. on Continuous Evaluation in Dairy Cattle*. College Park, MD. I. Misztal, ed. Dep. Anim. Sci., Univ. Illinois, Urbana.