NATIONAL COOPERATIVE DAIRY HERD IMPROVEMENT PROGRAM
HANDBOOK

A cooperative effort of:
● Dairy producers ● Breed associations ● AI organizations ● Cooperative Extension ● USDA—Extension Service — Agricultural Research Service

USDA-DHIA Animal Model Genetic Evaluations

G.R. Wiggans and P.M. VanRaden
Agricultural Research Service-USDA
Beltsville, Maryland

The United States Department of Agriculture (USDA) promotes genetic improvement of the national dairy cattle population by computing genetic evaluations of bulls and cows from data provided by Dairy Herd Improvement Associations (DHIA's) across the United States. Since July 1989, a statistical technique called the animal model has been used to compute these evaluations. The animal model predicts genetic merit of each animal in a population from the animal's own production records (if available) and the production records of all related animals.

The animal model replaced the Modified Contemporary Comparison (MCC) procedure used since 1974. Major benefits of the animal model are use of all relatives rather than just certain classes of relatives and use of exact statistical procedures (best linear unbiased prediction) rather than approximations. This factsheet describes how the animal model works, how it differs from MCC, and what benefits can be expected.[1]

Model

For genetic evaluation of dairy cattle, lactation records are described by a statistical model. This mathematical description includes 1) identification of the factors that contribute to the amount of milk produced in a particular lactation, 2) an in-

dication of how much of the variation between records is contributed by each factor, and 3) an indication of how closely factors are related to each other. The model is the blueprint for the evaluations. It determines how the data will be translated into rankings.

The USDA-DHIA animal model describes a cow's lactation record as the sum of the effects of her management group (m), genetic merit (animal effect, a), permanent environment (p), interaction of her herd and sire (c), and unexplained residual (e). If cow "kl" (daughter "l" of sire "k") had a lactation in management group "ij" (year-season, parity, and registry group "j" in herd "i"), her lactation yield (y) would be represented as:

$$y_{ijkl} = m_{ij} + a_{kl} + p_{kl} + c_{ik} + e_{ijkl}$$

The model does not include effects of age, length of lactation, and number of milkings per day, because lactation records are adjusted for these factors prior to genetic evaluation.

Management groups identify which lactations are compared with each other. They are determined by herd, month of calving, and lactation number. For Holsteins and Red & Whites, registry status also is considered. Initially, 2-month seasons are defined for first- and later-lactation groups. If a management group has fewer than five lactation records included, groups are combined in the following order: 2 months to 4 months, registered and grade together, 4 months to 6 months, first and later lactations together, 6 months to 12 months in steps of 2 months. A group with three or four lactation records is not combined

[1]Additional information included in: *Proceedings of the Animal Model Workshop.* 1988. **J. Dairy Sci.** 71:Suppl. 2.

Fact Sheet	Date	Page
H-2	1989	1

with another group if first and later lactations would have to be in the same group. Lactation records are not included if they cannot be compared with at least one other record. A management group composed only of daughters of a single sire does not contribute to that sire's evaluation.

An animal's genetic merit is the effect of all its genes (breeding value), not the effect of just the half that progeny receive (transmitting ability). Each offspring receives a different sample of genes from its parents, but the average genetic merit of progeny equals the average of parents' genetic merits. Animals that share the most genes provide the most information about each other's genetic merit. Expected fraction of genes that any two animals share is determined from pedigrees by a method published by USDA geneticist Sewall Wright in 1922.[2] Figure 1 provides an example.
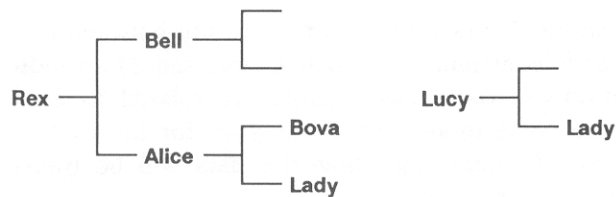
	Rex	Bell	Alice	Bova	Lady	Lucy
Rex	1	1/2	1/2	1/4	1/4	1/8
Bell	1/2	1	0	0	0	0
Alice	1/2	0	1	1/2	1/2	1/4
Bova	1/4	0	1/2	1	0	0
Lady	1/4	0	1/2	0	1	1/2
Lucy	1/8	0	1/4	0	1/2	1

Figure 1. Example pedigrees with expected fractions of genes in common.

Pedigrees are traced back as far as 1950 for the animal model. Every ancestral path eventually ends with unknown parents. Unknown parents are grouped, and their average merit is used in predictions for descendants. Groups are formed so that parents expected to have similar genetic merit are in the same group. Because of genetic improvement over time, more recent unknown parents

[2]Wright, S. 1922. *Coefficients of inbreeding and relationship.* **American Naturalist** 56:330.

have higher genetic merit. Therefore, unknown parents are grouped by birth year of their progeny, and several birth years may be included in the same group to insure that estimates of unknown-parent group effects are stable. Unknown parents also are grouped according to sex of the parent and sex of the animal itself. Separation by sex is necessary because the average merit of bulls (sires) usually is greater than that of cows (dams). For Holsteins, separate groups are defined for animals of U.S. and Canadian origin.

For most breeds, sires of cows as well as sires and dams of bulls usually are known. Lactation records of cows with an unknown sire are eliminated in initial editing. Parents without yield records and not related to at least two animals with yield records also are assigned to unknown-parent groups because they do not contribute any information to predictions. These unknown-parent groups are most important for grade animals because grades often are missing pedigree data.

The model includes two effects for each cow: genetic merit and permanent environment. The analysis is able to differentiate between these effects because the animal's genetic merit is correlated with its relatives' genetic merit, whereas its permanent environmental effect is assumed to be uncorrelated with those of its relatives. Effects of permanent environment, herd-sire interaction, and unexplained residual are assumed to be mutually independent and also independent of a cow's genetic ability. If cows are given special treatment because of high genetic evaluations, the assumption of independence will not be true, and subsequent genetic evaluations could be biased.

Heritability (h^2) is the proportion of differences between records due to genetics. For the animal model, h^2 has been set at .25, a compromise between the higher effective h^2 used for computing MCC sire evaluations and the lower h^2 for MCC cow evaluations. The environmental correlation between daughters of a sire in the same herd (the proportion of variation due to herd-sire interaction or c^2) is .14 as in the MCC. Accounting for c^2 limits the magnitude of an evaluation for bulls with daughters in only a few herds or unequal daughter distribution among herds. The correlation between repeated records of the same cow (repeatability or r) has been set at .55 as compared with .50 for the MCC. This repeatability is the sum of h^2, c^2, and also p^2 (.16), the proportion of variation due to permanent environment. The same h^2, c^2, and p^2 are used for milk, fat, and protein.

Data

Data for animal model evaluations include lactation yield information (milk, fat, and protein) and pedigree information. Lactations back to 1960 calvings and pedigree data back to 1950 are included. Ancestors must be evaluated to account for selection. Lactation records for cows with a missing first-lactation record are excluded from evaluations that affect relatives to reduce selection bias. A bias could occur if daughters with missing first-lactation records were genetically superior to the average of the bull's daughters as the result of culling on first-lactation performance. Lactations after fifth are excluded because of their reduced value in estimating genetic merit. Relatively few animals in a herd have more than five lactations; therefore, few contemporaries of similar age are available for valid comparisons. In addition, the influence of environmental effects on the record increases with each lactation.

Solutions

Solutions are obtained by a process of repeated calculations called iteration. Initially, all estimates of model effects are 0 or their value from the previous evaluations. Values computed in the first round of iteration are used in the second round; those from the second in the third; etc. Iteration continues until the differences in solutions between rounds becomes acceptably small. Iteration allows the contribution of each animal to be passed on to all its relatives.

Management group effect is estimated as the weighted average of differences between lactation yield and other effects in the model over all lactations in the management group:

$$\hat{m} = \Sigma[w_{len}(y - \hat{a} - \hat{p} - \hat{c})]/\Sigma w_{len}$$

where a $\hat{}$ indicates an estimate of the effect, Σ indicates summation, and the lactation length weights (w_{len}) depend on number of days in milk, type of test, and parity.

Permanent environmental effect is predicted as the weighted sum of the differences between a cow's lactations and the other effects in the model divided by the sum of the weights plus a variance ratio, $(1-r)/p^2 = 2.8$, that tends to reduce the magnitude of the estimate:

$$\hat{p} = \Sigma[w_{len}(y - \hat{m} - \hat{a} - \hat{c})]/(\Sigma w_{len} + 2.8)$$

Herd-sire interaction effect is predicted as the weighted sum of differences between lactation yield and the other effects in the model for all the lactations of all a bull's daughters in a herd divided by the sum of the weights plus a variance ratio, $(1-r)/c^2 = 3.2$:

$$\hat{c} = \Sigma[w_{len}(y - \hat{m} - \hat{a} - \hat{p})]/(\Sigma w_{len} + 3.2)$$

An animal's predicted genetic merit ($\hat{a}$) is computed as a weighted combination of three sources of information: 1) average of its sire's ($\hat{a}_s$) and dam's ($\hat{a}_d$) predicted merits, 2) its yield deviation (YD) where:

$$YD = \Sigma[w_{len}(y - \hat{m} - \hat{p} - \hat{c})]/\Sigma w_{len}$$

and 3) average of contributions from progeny. A progeny contribution is twice the progeny's predicted merit ($\hat{a}_p$) minus the mate's predicted merit ($\hat{a}_m$); estimates of unknown-parent group effects are substituted for unknown parents or mates. Mathematically, $\hat{a}$ can be represented as:

$$\hat{a} = w_1[(\hat{a}_s + \hat{a}_d)/2] + w_2(YD) + w_3(\overline{2\hat{a}_p - \hat{a}_m})$$

where the w's are weighting factors in fractional form that sum to 1. For w_1, the numerator is 2 if both parents are evaluated, 4/3 if only one parent is evaluated, or 1 if neither parent is evaluated. For w_2, the numerator is $(\Sigma w_{len})[h^2/(1-r)]$. For w_3, the numerator is half the number of progeny, but progeny of unknown mates count only 2/3. The three w's have the same denominator, which is the sum of their numerators. Predicted merit for animals without records (for example, bulls) is computed the same way except that w_2 is 0.

Although only parents and progeny appear to be included in an animal's predicted genetic merit, all relatives do contribute. Information from more distant relatives is included through the animal's parents and progeny because the evaluation of each parent or progeny includes its parents and progeny. This is an application of the method developed by Dr. C.R. Henderson for including all relatives when inbreeding is ignored.[3]

Predictions for the Holstein and Red & White breeds are computed jointly so that the many relationships across the breeds can be included and the predictions compared across breed.

[3]Henderson, C.R. 1976. *A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values.* **Biometrics** 32:69.

Fact Sheet	Date	Page
H-2	1989	3

Computation of Final Evaluations

To compute final evaluations expressed as transmitting ability, predictions of genetic merit for each breed first are adjusted so that cows born in 1985 average 0. The YD is adjusted by the same amount. This adjustment imposes a genetic base, a reference point for comparison of animals. The base is labeled by breed and year that the base was changed. In anticipation of future base changes on years evenly divisible by 5, the genetic base for the first implementation of the animal was designated as "90" rather than "89." For example, A90 denotes the genetic base for Ayrshires, for which the average evaluation of Ayrshire cows born 5 years previously (1985) is set to 0. For the combined evaluations of Holsteins and Red & Whites, the base is labeled "HW."

Adjusted predicted genetic merit (breeding value) is divided by 2 to obtain predicted transmitting ability (PTA):

$$PTA = \hat{a}/2$$
$$= w_1(PA) + w_2(YD/2) + w_3(\overline{2PTA_p - PTA_m})$$

where $PA = (PTA_s + PTA_d)/2$, the average transmitting ability of the sire (PTA_s) and the dam (PTA_d) or parent average; PTA_p is PTA of a progeny; and PTA_m is PTA of a mate (the progeny's other parent). This formula differs from the formula for $\hat{a}$ in that YD is divided by 2. For the other terms, division by 2 is not necessary because they already are expressed as transmitting abilities. The term PTA is used for both cows and bulls. Comparison of PA with PTA indicates the impact of progeny and records on an animal's evaluation.

Table 1. Averages of standardized yield traits for cows born in 1985 by breed.

Breed	Milk	Fat	Fat	Protein	Protein
	(lb)	(lb)	(%)	(lb)	(%)
Ayrshire	13,848	535	3.86	454	3.28
Brown Swiss	15,907	613	3.85	550	3.45
Guernsey	12,715	574	4.51	441	3.47
Holstein/ Red & White	19,004	681	3.58	589	3.10
Jersey	12,855	606	4.71	473	3.68
Milking Shorthorn	13,673	493	3.61	448	3.27

The PTA's for fat and protein percentages are derived from yield evaluations combined with first-lactation, mature-equivalent yields of cows born in 1985 for the appropriate breed (Table 1). The PTA for protein yield is calculated as a function of the PTA for protein percentage (computed only from records with protein information) and PTA for milk yield computed from all records so that PTA's for milk and protein yields have a similar basis. If an animal does not have protein information, its PA for protein percentage is used to estimate a PTA for protein yield. Frequently, if an animal does not have a protein evaluation, neither does its dam; therefore, the PA for protein percentage would include an unknown-parent group estimate. For bulls, this process is particularly important because the subset of daughters with protein information could be quite different from his complete set of daughters.

The requirement for first-lactation data is relaxed for protein evaluations. The problem of selection bias is not expected to be as great for protein as for milk and fat. Widespread protein testing was introduced relatively recently; therefore, selection emphasis on protein has been less. In addition, cows are required to have first-lactation milk and fat data for their protein data to be included in computing evaluations for protein percentage.

Economic indexes called PTA dollars (PTA$) are computed from PTA's. Separate PTA$ are calculated for milk and fat; milk, fat, and protein; and cheese yield. Percentiles are based on the PTA$ that includes milk, fat, and protein. For bulls, rankings are based on bulls that were in active artificial-insemination service following the previous evaluation. For cows, rankings are based on cows with recent lactations.

Predicted producing ability (PPA) includes predictions of c, p, and a. The PPA minus twice PTA is the sum of estimates for c and p. Thus, PPA is useful as an indicator of future production of a cow and for determining estimates of other effects in the model.

Daughter yield deviation (DYD) is the weighted average YD of a bull's daughters adjusted for merit of their dams. This adjustment for merit of mates is not included in YD. The DYD provides an indication of the performance of the bull's daughters without consideration of his parents or sons. The animal model's YD is similar to MCC's Modified Contemporary Deviation.

Table 2 lists animal model information that is distributed to the dairy industry.

Table 2. Information generated from the USDA-DHIA animal model.

Animal model information	Description
Predicted transmitting ability (PTA)	One-half breeding value; adjusted so that cows born in 1985 average 0
Average standardized yield	Averaged over lactations for cows and over daughters for bulls
Predicted producing ability (PPA)	Prediction of a cow's performance in future lactations; sum of predictions of breeding value, permanent environmental, and herd-sire interaction effects
Yield deviation (YD)	Weighted average yield adjusted for management group, permanent environmental, and herd-sire interaction effects
Daughter yield deviation (DYD)	Weighted average of YD's of a bull's daughters adjusted for merit of their dams
Parent average (PA)	Average PTA of parents; if either parent is unknown, unknown-parent group effect is substituted
Reliability (REL)	A measure of amount of information in the evaluation; same value for milk and fat PTA
REL_{PA}	Amount of information in PA; calculated as one-fourth of sum of parents' REL's
PTA dollars (PTA$)	Economic index combining evaluations for milk and components weighted by product value; calculated as for Modified Contemporary Comparison
Percentile	Ranking based on PTA$ that includes milk, fat, and protein

Indication of Accuracy

The measure of amount of information in an animal's evaluation is called reliability (REL). The method for computing REL is an extension of the MCC procedure for Repeatability (RPT). In addition to sources of information used in MCC RPT, REL includes contributions from parents and sons for bulls and from progeny for cows. Unknown-parent groups do not contribute to REL.

Including progeny for cows means that cows producing many progeny through embryo transfer now can attain high REL. The name was changed because (1) REL generally is higher than RPT as a result of contributions from additional relatives and (2) confusion existed between "Repeatability" for measuring accuracy and "repeatability" for defining similarity between repeated records.

For an animal with no records or progeny information, REL is one-fourth the sum of parent REL's, which also is REL of PA.

For animals with more information available than just that from parents, REL is computed from daughter equivalents. Daughter equivalents provide a common unit for measuring the amount of information contributed by an animal's parents, its own records, and its progeny.

The amount of information that a sire receives from any one herd is limited because the model includes an effect for interaction of herd and sire. Table 3 shows an example of number of daughter equivalents that would be contributed to a sire by daughters in the same herd. This example assumes that each daughter has one record, a dam with a known PTA, and a large number of management group mates that does not include paternal half-sibs. For these conditions, daughter equivalents contributed to the sire are calculated as $1/[.16 + (.84/d)]$, where d is the number of sire's daughters within the herd. If any daughter has more than one record, d is replaced by the sum of $1/[.39 + (.61/\Sigma w_{len})]$.

Table 3. Example daughter equivalents contributed to sire by daughters in the same herd.

Number of daughters in herd	Daughter equivalents contributed to sire	Daughter equivalents per daughter
1	1.0	1.00
2	1.6	.81
5	2.9	.58
10	4.0	.40
25	5.1	.20
50	5.6	.11
100	5.9	.06

Table 4 shows example daughter equivalents contributed to a cow by various relatives. As REL of a relative's evaluation increases, so do the number of daughter equivalents that it contributes. An animal's REL can be calculated by summing the daughter equivalents (n) from all sources (parents, own records, and progeny) and then applying the formula REL = $n/(n+14)$.

Comparison of Animal Model and MCC Procedures

The animal model evaluation system has similarities to MCC. Later-lactation records are included as is a herd-sire interaction effect. Provision is made for comparing a lactation with appropriate first or later lactations of contemporaries. The method for computing the measure of the amount of information in animal model evaluations is an extension of the MCC method for RPT; however, contributions from more relatives are included in the animal model, and the name has been changed to REL.

Differences between the two methods are in Table 5. The primary difference is that the animal model uses additional information and employs many rounds of iteration to improve accuracy and to insure that information from each animal is included in evaluations of all its relatives.

The changes between the MCC and animal model genetic bases are in Table 6 by breed and yield trait.

Table 4. Example daughter equivalents contributed to cow reliability (REL) by various sources of information.

Relative	Information available	Daughter equivalents
Parents	Sire with 70% REL and dam with 30% REL	4.7
	Sire with 99% REL and dam with 50% REL	8.3
	Sire with 99% REL and dam with 99% REL	14.0
Self	1 lactation record	4.7
	3 lactation records	7.8
	5 lactation records	9.0
Daughter	1 lactation record	1.0
	3 lactation records	1.5
	5 lactation records	1.7
Son	1 daughter with 1 lactation record	.2
	10 daughters in 10 herds, each with 1 lactation	1.8
	50 daughters in 50 herds, each with 1 lactation	4.4
	100 daughters in 100 herds, each with 1 lactation	5.4
	Evaluation with 99% REL	7.0

Table 5. Differences between animal model (AM) and Modified Contemporary Comparison (MCC) evaluations.

Characteristic	AM	MCC
Animals evaluated	All (simultaneously)	Recent only (bulls, then cows)
Merit of mates considered	Yes	No
Dams contribute to sons	Yes	No (ancestor merit includes maternal grandsire)
Sons contribute to parents	Yes	No
Daughters contribute to dams	Yes	No
Base definition	Cows born in 1985	Bulls weighted by number of daughters first calving in 1982
Environmental group definition	Management group (registered-grade, 2 mo, first-later lactation, groups combined to include 5 lactations, cow included)	Contemporary group (5 mo, centered, some contribution of later lactations to first, cow excluded)
First lactation required	Yes (cows without first lactation records evaluated separately)	No (cows without first lactation records receive less weight)
Lactations included	1-5	1-15
Later herd lactations included	Yes (in supplemental evaluation)	Yes
Reliability components:		
Parents for males	Yes	No
Daughters for females	Yes	No
Sons	Yes	No

Supplemental Evaluation

Lactation records for cows without a first-lactation record are not included in the main evaluations so that selection bias can be minimized in relatives' evaluations. However, these records are used in calculating a supplemental evaluation for the cows so that they can have the most accurate evaluation possible (unless the records are unrepresentative, perhaps because of preferential treatment). A cow with a missing first-lactation record may also have a main evaluation if she has progeny, but this evaluation would not include any of her lactation records.

For supplemental evaluations, a, p, and c are computed with management group estimates and predicted genetic merit of relatives from the main evaluation. For cows that change herds, supplemental evaluations also are computed to calculate a common p across herds. A common p is not predicted in the main evaluation because of computational complexity.

Table 6. Changes between the genetic bases for the Modified Contemporary Comparison (PD82) and the animal model.

Breed	Milk	Fat	Protein
	(lb)	(lb)	(lb)
Ayrshire	89	1.1	.4
Brown Swiss	170	8.3	3.7
Guernsey	196	7.7	3.0
Holstein	200	6.8	3.1
Jersey	224	10.0	3.7
Milking Shorthorn	352	10.3	12.1
Red & White	742	25.6	19.8

Supplemental evaluations are restricted to cows born in the preceding 10 years (that is, those that might add lactation records). Older cows still being milked probably would have exceeded the five-lactation limit and, therefore, would not add information.

Summary

The animal model allows simultaneous genetic evaluation of bulls and cows with all relationships included. Previously computing constraints limited this approach to evaluation within herd. Recent advances in animal breeding theory and increased computer capacity made animal model evaluation computationally feasible for national data sets.

The USDA's Animal Improvement Programs Laboratory developed an implementation of the animal model that promises to improve the accuracy of evaluations of U.S. dairy cattle. Supplemental information is provided to assist in tracing the source of an individual evaluation.

The REL reflects amount of information included in an animal's PTA but not the quality of that information. No known system or measure of accuracy can account for manipulation or misrepresentation. Computers can aid greatly in breeding decisions, but subjective judgment on credibility of original data still is required. For data that follow the assumptions of the model, evaluations computed with the animal model offer the best predictions of future performance.