Including Information from Records in Later Herds in Animal Model Evaluations

G. R. WIGGANS and P. M. VANRADEN
Animal Improvement Programs Laboratory
Agricultural Research Service
United States Department of Agriculture
Beltsville, MD 20705

ABSTRACT

Computing strategies were developed to allow information from later herds to be included in animal model evaluations for cows that changed herds. Records from later herds were modeled as containing the same permanent environment and herd-sire interaction as did the cow's records from the first herd. This allowed later herd records to contribute to estimation of these effects and made repeatability the same across herds as within herd. The national evaluation system was changed to store management group deviations from later herds and include them when processing information from the first herd. This method of including all lactation records for cows that changed herds was tested on United States data for Ayrshires, Brown Swiss, Guernseys, and Jerseys. Correlations between parent average and yield deviation generally increased slightly for daughters of cows with lactations in more than one herd compared with the corresponding correlations if lactations from later herds were excluded. For January 1990 USDA-DHIA genetic evaluations, records from later herds were included for all breeds, which eliminated the need for two evaluations for cows that changed herds (the main evaluation with data from only the first herd and a supplemental evaluation with all lactations through fifth included).
(**Key words**: animal model, genetic evaluation, later herd)

Received February 5, 1990.
Accepted May 9, 1990.

INTRODUCTION

Adoption of an animal model for genetic evaluation of yield traits of dairy cattle in the United States required several changes in which records were included in evaluations (6, 7). Records from cows without a reported first lactation were excluded because of possible selection bias. For cows that changed herds, only data from the first herd were included. A cow's records from later herds were excluded to simplify implementation of the animal model system and to avoid possible bias from adopting the strategy used for dairy goats. For dairy goats, a separate permanent environmental effect was estimated for each herd in which a doe had lactations (5). Similarly, herd-sire interaction effects were estimated from each herd in which lactations were produced. This model was not appropriate for dairy cows because cows that change herd may be those most likely to have large permanent environmental effects. If a new permanent effect is estimated for each herd, estimates of permanent environmental effects are more severely regressed toward 0, and genetic estimates may be inflated. In addition, repeatability within herd is not maintained across herds.

A supplemental evaluation system had been developed for cows without a reported first lactation. This system was adapted to include all lactations through fifth for cows that changed herds (6, 7). This procedure was unsatisfactory, because the two evaluations for one cow led to confusion about which evaluation was best. Also, supplemental evaluations were computed only for cows born in the preceding 10 yr, and substantial change could occur in a cow's evaluation when the cow first reaches the 10-yr limit.

The goal of this research was to develop a method to include a cow's records from later herds in the main evaluation and to determine if such evaluations are more accurate in predicting progeny performance than evaluations that do not include records from later herds.

MATERIALS AND METHODS

Model

The model assumed for July 1989 evaluations (6) was modified to include lactation information from later herds. The revised model implemented in January 1990 was

$$y_{ijkl} = m_{ij} + a_{kl} + p_{kl} + c_{i'k} + e_{ijkl}$$

where y_{ijkl} = milk, fat, or protein yield of cow kl (daughter l of sire k) in herd i in year-season, parity, and registry group j; m, a, p, c, and e are fixed management and random animal, permanent environmental, herd-sire interaction, and residual effects; and i′ indicates the herd in which the first lactation was produced. The herd-sire effect from the first herd was assumed for all the cow's herds; consequently, repeatability is constant across herds. Constant repeatability also was assumed by Modified Contemporary Comparison procedures (2, 3). Animal effect was breeding value and included effects for unknown-parent groups. Variance components scaled to a phenotypic variance of 1 were .25 for genetic, .16 for permanent environmental, .14 for herd-sire, and .45 for residual variances, which resulted in heritability of .25 and repeatability of .55. The same variance components were used for milk, fat, and protein. The (co)variance matrix of residuals was diagonal; elements of its inverse were lactation weights determined by lactation length (6). Changes to the national animal model evaluation system that were necessary to include later herd data follow.

Computational Steps

Data Selection. A first lactation record continued to be required, but the requirement that the first lactation have management group mates was dropped. Previously, if a cow did not have management group mates for her first lactation, all her lactation records were excluded, and this in turn caused other cows to lose their management group mates. Later records of cows with no management group mates for first lactation should not cause selection bias, because any previous selection on records with no management group mates probably was ineffective. Lactation records without management group mates provide no information and were excluded; however, later lactations of such cows were included. Indicators were added to identify cows with a first lactation reported and cows with protein information for at least one of their first five lactations.

A file including the identification (ID) of cows with data in later herds was created. An indicator was added to the file containing yield data to identify if the cow's first lactation was produced in that herd. Every cow was represented in her first herd even if none of the lactations in that herd had management group mates. This ensured that there would be a first herd for every cow to trigger estimation of herd-sire interaction and permanent environmental effects. The same strategy was necessary for protein, because a cow with protein information in later herds might not have protein information from her first herd.

Cows that start their first lactation with an abortion presently are classified as not having a first lactation (4). When data editing programs can be modified, such records will be retained to indicate that later lactation records of such cows can be included, because there was no effective selection based on first lactation.

Preparation for Iteration. The ID of cows with data in later herds was stored in a hash table (6), which is an array in memory, to allow rapid checking for presence of later herd data. As yield data were processed, the array was checked for the cow's ID. If her ID was not found, a 0 was stored in her record; otherwise, the subscript of the array was stored and the equation number for her herd-sire interaction effect stored in a vector if the herd was the cow's first herd. With only one permanent environmental effect per cow, the equation number for her animal effect also was used as the equation number for her permanent environmental effect.

Iteration. New estimates of permanent environmental and herd-sire interaction effects always were computed when first-herd data were encountered. For the first iteration, this meant that herd-sire interaction and permanent environmental effects were estimated with data accumulated through only the first herd. The sum of lactation weights for the cow also was accumulated during the first iteration and, therefore, corresponded with management group deviations (y − m̂) collected. The accumulator for management group deviation was set to 0 after processing the first herd, not at the end of the round. Subsequent rounds had herd-sire interaction and permanent environmental effects computed from management group deviations from all of the cow's lactations, because deviations were accumulated from herds following the first in the preceding round and from herds through the first in the current round. This strategy enabled continued estimation of management, herd-sire interaction, and permanent environmental effects within herd, which in turn saved memory and increased convergence rate by allowing right-hand sides for breeding value to include more current values for these effects.

Determination of the equation number for herd-sire interaction effect was complicated by the presence of cows in later herds. These cows did not signal the start of a new herd-sire group. Sires that had only one daughter and no known ancestors were treated as unknown parents. Each instance of a sire assigned to an unknown-parent group was given a separate interaction effect.

Reliability. Records from later herds that were processed after a daughter's first herd did not contribute to the sire's reliability (measure of accuracy). This restriction enabled bull information to be collected during one pass of the data. The number of herds continued to be a count of first herds. If a cow had no usable lactation records in her first herd or only paternal half-sibs as management group mates, she did not contribute to herd count. Cow reliability included information from all herds.

A change implemented with this revision of the system was to evaluate all dams of young bulls being progeny tested even if the dam had no lactations. Bulls resulting from juvenile embryo transfer may be available for progeny test before their dams generate lactation records. Inclusion of these dams forced calculation of their PTA, which were then available for computing parent averages of their sons. Previously, these dams had been represented as unknown parents. This revision had no effect on accuracy of evaluations except that averages of parents' PTA (parent averages) reported for some young bulls were more accurate.

TABLE 1. Distribution of cows with lactation records in more than one herd by breed.

Breed	Cows with records in more than one herd	Portion of cows evaluated
	(No.)	(%)
Ayrshire	13,382	10.1
Brown Swiss	17,668	11.3
Guernsey	45,371	9.0
Holstein[1]	681,583	6.9
Jersey	67,477	10.7

[1]Holstein data from January 1990 evaluations.

System Test

Data for July 1989 USDA-DHIA animal model evaluations of Ayrshires, Brown Swiss, Guernseys, and Jerseys were included in evaluations calculated with the later herd system. Cows with lactations in more than one herd were identified. Distribution of cows is in Table 1 by breed. Their daughters' evaluations were obtained from July 1989 evaluations. Daughters were required to have a birth date of 1970 or later and to have management group mates for their first lactations.

Effect of including data from later herds in the main evaluation was measured by calculating correlations between parent average and yield deviation of daughters of cows with lactations in multiple herds and comparing correlations from the later herd system with correlations from July 1989 evaluations. The July 1989 yield deviations for daughters were from the evaluation distributed to the industry that was labeled "evaluation used for relatives". Effect of genetic trend on correlations was removed by computing correlations on deviations within birth year.

RESULTS AND DISCUSSION

Including a cow's records from later herds increased memory requirements, because vectors were added to indicate presence of records in later herds, store management group deviations from later herds, and store herd-sire equation numbers. Number of cows evaluated increased, because cows with later lactations but no management group mates for first lactation were included and because lactations in later herds provided management group mates for other cows that formerly had no group mates.

TABLE 2. Correlations within year of parent average[1] and yield deviation for milk and fat based on July 1989 USDA-DHIA animal model evaluations excluding and including a cow's lactation records from later herds.

		Milk		Fat	
Breed	Number of daughters	Later herds excluded	Later herds included	Later herds excluded	Later herds included
Ayrshire	5953	.575	.578	.573	.580
Brown Swiss	9966	.579	.582	.581	.583
Guernsey	17,727	.574	.571	.572	.571
Jersey	40,694	.632	.634	.600	.601

[1]Average of parent PTA.

The memory requirement of the iteration program was made flexible by allowing data in excess of memory capacity to be read from disk each iteration. For January 1990 Holstein evaluations, about 30% of yield data were obtained from disk each iteration.

Correlations from July 1989 evaluations and from the system including later herds are in Table 2. Although correlations were similar, correlations from the later herd system were higher for both milk and fat for all breeds except Guernsey. A larger increase in correlations was expected because later herds provide additional data, which should improve accuracy of evaluations. Because cows in later herds can serve as management group mates, additional management group mates or more homogeneous management groups may provide an improvement to accuracy for animals not included in Table 2 (cows with lactation records in only one herd). Variation among breeds in the ratio of number of daughters to number of cows was a result of differences in trends in population size and, therefore, the number of daughters reaching milking age.

A risk with this system is that purchased cows may not receive the same treatment as their herdmates. Records from later herds may be biased, especially for bull dams. Financial incentives can cause preferential treatment of certain cows, but moving all bull dams to central embryo transfer facilities could result in more similar management and more accurate evaluations (1). Decisions on individual cows must consider whether the cow and her management group mates actually were managed similarly.

The major benefit of the later herd system is that it provides evaluations to producers that include all data. The ability to use purchased cows as management group mates and to have only one evaluation for cows that change herds rather than two simplifies explanation of the evaluation system and facilitates understanding of animal model procedures.

ACKNOWLEDGMENTS

This research was conducted using the Cornell National Supercomputer Facility, a resource of the Cornell Theory Center, which is funded in part by the National Science Foundation, New York State, IBM Corporation and members of the Center's Corporate Research Institute. Assistance of S. M. Hubbard in manuscript review is acknowledged.

REFERENCES

1 Bierma, J. 1990. New breeding program in West Germany boosts the role of individual cows. Holstein World 87:116.

2 Dickinson, F. N., H. D. Norman, R. L. Powell, L. G. Waite, and B. T. McDaniel. 1976. Procedures used to calculate the USDA-DHIA Modified Contemporary Comparison. Page 18 *in* USDA Prod. Res. Rep. No. 165, Washington, DC.

3 Powell, R. L., H. D. Norman, and F. N. Dickinson. 1976. The USDA-DHIA Modified Contemporary Comparison Cow Index. Page 35 *in* USDA Prod. Res. Rep. No. 165, Washington, DC.

4 Wiggans, G. R. 1985. Codes for transfer of NCDHIP data. Natl. Coop. Dairy Herd Improvement Progr. Handbook, Fact Sheet H-11, Washington, DC.

5 Wiggans, G. R. 1989. Animal model evaluations of dairy goats for milk, fat, and protein yields with crossbred animals included. J. Dairy Sci. 72:2411.

6 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.

7 Wiggans, G. R., and P. M. VanRaden. 1989. USDA-DHIA animal model genetic evaluations. Natl. Coop. Dairy Herd Improvement Progr. Handbook, Fact Sheet H-2, Washington, DC.