

**THE EVALUATION OF DAIRY SIRE PROGENY TESTS BY THE UNITED STATES
DEPARTMENT OF AGRICULTURE**

**Evaluación de las pruebas de descendencia en toros de razas lecheras
en los Estados Unidos**

**L'évaluation des épreuves de descendance chez les taureaux de races
laitières aux Etats-Unis**

F. N. DICKINSON *

Since 1967, the United States Department of Agriculture has used the herdmate comparison method to calculate dairy sire genetic evaluations. This method is based on the assumptions that all herdmates are drawn randomly from one genetically homogeneous population, that there is no genetic trend in this population, and that bulls' daughters and their herdmates are subjected to the same severity of culling for milk and milk fat yield. These assumptions are no longer valid, primarily because of the success in identifying genetically superior bulls for use in artificial insemination in the United States since the wide use of the herdmate comparison method. Therefore, research has been conducted for several years on Sire Summary methods that will eliminate these assumptions and permit the accurate genetic evaluation of all dairy bulls. Two separate methods are being developed, and computing systems have been designed and programmed for each.

The first of these methods is called the *USDA-DHIA Modified Contemporary Comparison*. It differs markedly from *Contemporary Comparisons* used in most other countries. All available lactation records are used for both daughters and herdmates. Daughter-herdmate deviations are weighted so that those of similar parity receive greater emphasis than those of different parities; hence the name Modified Contemporary Comparison. Each daughter-herdmate deviation is adjusted for the genetic merit of the herdmate's sire.

The modified contemporary deviations are iterated so that the genetic evaluations become more accurate with repeated summaries. The iterated modified contemporary deviations are regressed to the genetic group to which the bull belongs rather than to the overall breed average. This genetic grouping procedure

* United States Department of Agriculture, USDA; Agricultural Research Service, Animal Physiology & Genetics Institute, Animal Improvement Programs Laboratory, Bldg. 263, ARC-EAST, Beltsville, Maryland 20705, United States.

is based on a pedigree estimate of transmitting ability for each bull, based on his sire's and maternal grandsire's previously calculated Predicted Differences. Thus far, no one has been able to write a mathematical model to represent the *Modified Contemporary Comparison*. This should not be considered a serious disadvantage of the method because the present *USDA-DHIA Herdmate Comparison Method* has been used effectively for several years and, to the best of my knowledge, a mathematical model has not been written to represent it. Research with empirical data has shown that the adjustment for the genetic level of the herdmate sires,

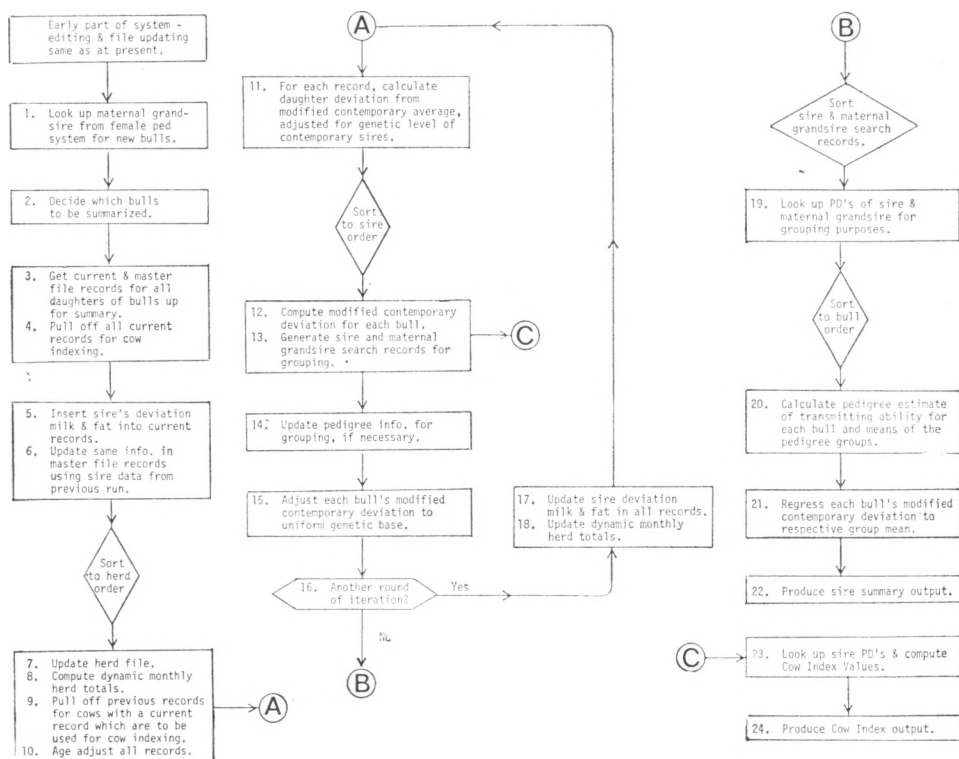


FIG. 1.

accomplished by repeated solutions, should eliminate bias due to genetic trend as well as that due to the genetic merit of the herdmates with which each bull's daughters are compared.

A flowchart showing the major computing steps of the Modified Contemporary Comparison Method is shown in Figure 1. Data are passed to this part of the system from the editing and file updating procedures in the early part of the system. Step 1 is to look up the sire and maternal grandsire of any new bulls to be summarized. Sire and maternal grandsire information is transferred to

the master sire file for new bulls entering the system for the first time. Step 2 is to decide which bulls are to be summarized in the current run. This is determined by a set of rules based on Repeatability of the Predicted Difference (P.D.) and the amount of new information.

After the bulls to be summarized are determined, the records coming in to this run for the first time (called current records) plus records on those bull's daughters from the master file are pulled off (step 3). In addition, all current records are pulled off for later use in cow indexing, regardless of whether the cows' sires are up for summary or not (step 4). Next the sire's deviation milk and fat is inserted into all the current records (step 5). The sire's deviation milk and fat is also inserted into all records from the master file; data from the previous Sire Summary run (step 6) are used. These records are then sorted to herd order.

In this procedure, a herd file must be maintained. This file is updated at each run with new records entering the system (step 7). After this, dynamic monthly herd totals are computed (step 8) as required for the rolling herdmate average procedure.

Also, previous records are pulled off for cows that have a current record and are to be cow indexed but whose sires are not up for summary (step 9).

All the records that are to be used for Sire Summaries or Cow Indexes are now available. So at this point, all records are standardized for age and month of calving (step 10). Records to be used must be adjusted each run because all records stored in USDA master files are unadjusted. At this point in the system, point A, the iterative Sire Summary calculations are begun. First, the modified contemporary average adjusted for the genetic level of the contemporary sires is calculated for each daughter record (step 11). These records are sorted to sire sequence. The modified contemporary deviation for each bull is calculated (step 12), and sire and maternal grandsire search records, which will be needed to look up PD's for pedigree grouping, are generated (step 13). The lactation records that were used in step 12 are now available for cow indexing at point C. On the first round of iteration, the pedigree information, to be used later for grouping, is updated (step 14); *i. e.*, missing ID information is filled in and ID errors are corrected. Next, each bull's modified contemporary deviation is adjusted to the uniform genetic base so that summaries can be directly compared from one run to the next (step 15).

At this point (step 16), decide whether this is the final round of iteration. If it is the final round of iteration, then the system proceeds to point B. If it is not, the sire deviations for milk and fat in all records (step 17) and in the dynamic monthly herd totals (step 18) must be updated and another round of iteration begun at A.

The sire and maternal grandsire search records generated at step 13 are sorted, and the PD's of the appropriate sires and maternal grandsires are looked up (step 19). These search records are then sorted back to bull order. A pedigree estimate of transmitting ability for milk yield for each bull and the means of the pedigree groups are calculated (step 20). Each bull's modified contemporary deviation is regressed to that bull's pedigree-group mean (step 21). The genetic evaluations (Predicted Differences) on the bulls summarized are ready at this

point to be printed and distributed to the dairy industry. This is done in (step 22).

The final procedure, at point C, is to look up the sire PD's and compute the *Cow Index Values* (step 23). The final step is to print and distribute the Cow Indexes (step 24). These then are the major steps in calculating the USDA-DHIA Modified Contemporary Comparison Method.

The second Sire Summary method is the Linear Model Method, which has been developed cooperatively with Dr. C. R. HENDERSON. The mathematical model for this method is:

$$y_{fijklm} = g_i + s_{ij} + h_{kl} + (s \times h)_{fijk} + c_{fikm} + e_{fijklm}$$

where f is the category of sire (artificial insemination or natural service) for computational purposes only, i is the sire group, j is the sire within the i^{th} group and f^{th} category, k is the herd, l is the year season within the k^{th} herd, m is the cow within the fij^{th} sire and k^{th} herd, y is a lactation record adjusted for age and season of calving, g is the sire group effect (fixed), s is the sire effect (random), h is the herd-year-season effect (fixed), c is the cow effect (random), $s \times h$ is the environmental sire \times herd effect (random), and e is the random residual effect. Thus, this model considers sources of variation due to sire groups (G), herd-year-seasons (HYS), cows and the environmental correlation among a bull's daughters in each herd ($S \times H$). Of these, G and HYS are considered fixed and the rest are random sources of variation. With this model, both AI and natural service (NS) bulls can be summarized by using all available records on the daughters. This model also accounts for the intra-herd correlation among a bull's daughters, which is very important to consider when NS bulls in the United States are summarized.

A flowchart of the major steps in the solution of this model is shown in Figure 2. Data are passed to this part of the system from the editing and file updating procedures that constitute the early part of the present *USDA-DHIA Sire Summary System*. The lactation records are sorted as indicated at the top of Figure 2 and passed to the program A. This program performs 10 functions as shown. Steps 1 to 6 are performed sire by sire as follows. The individual cow equations are generated, and each equation is written on tape for later use in cow indexing (steps 1 and 2). Then the cow equations are absorbed while the $S \times H$ and the HYS equations for each sire are accumulated (step 3). Each $S \times H$ equation is written out after all cow records for a particular sire and herd have been absorbed (step 4). Then the $S \times H$ equation is also absorbed (step 5). If there are any natural service sires (NSS) in that herd, they are absorbed (step 6). Each sire is treated as if it occurred in only that one herd. The absorption of sires is not necessarily limited to only natural service sires. However, the primary purpose for the sire absorption procedure is to eliminate the large number of natural service sires in order to reduce the sire equations in the initial solution to a manageable number.

When the end of a herd is reached, steps 7 to 10 are performed. First, the inverse of the $HYS \times HYS$ submatrix (for the herd just concluded) is obtained (step 7). This inverse is multiplied by the appropriate right-hand members, and the product is written out on the same tape as the $S \times H$ equations (step 8). Then

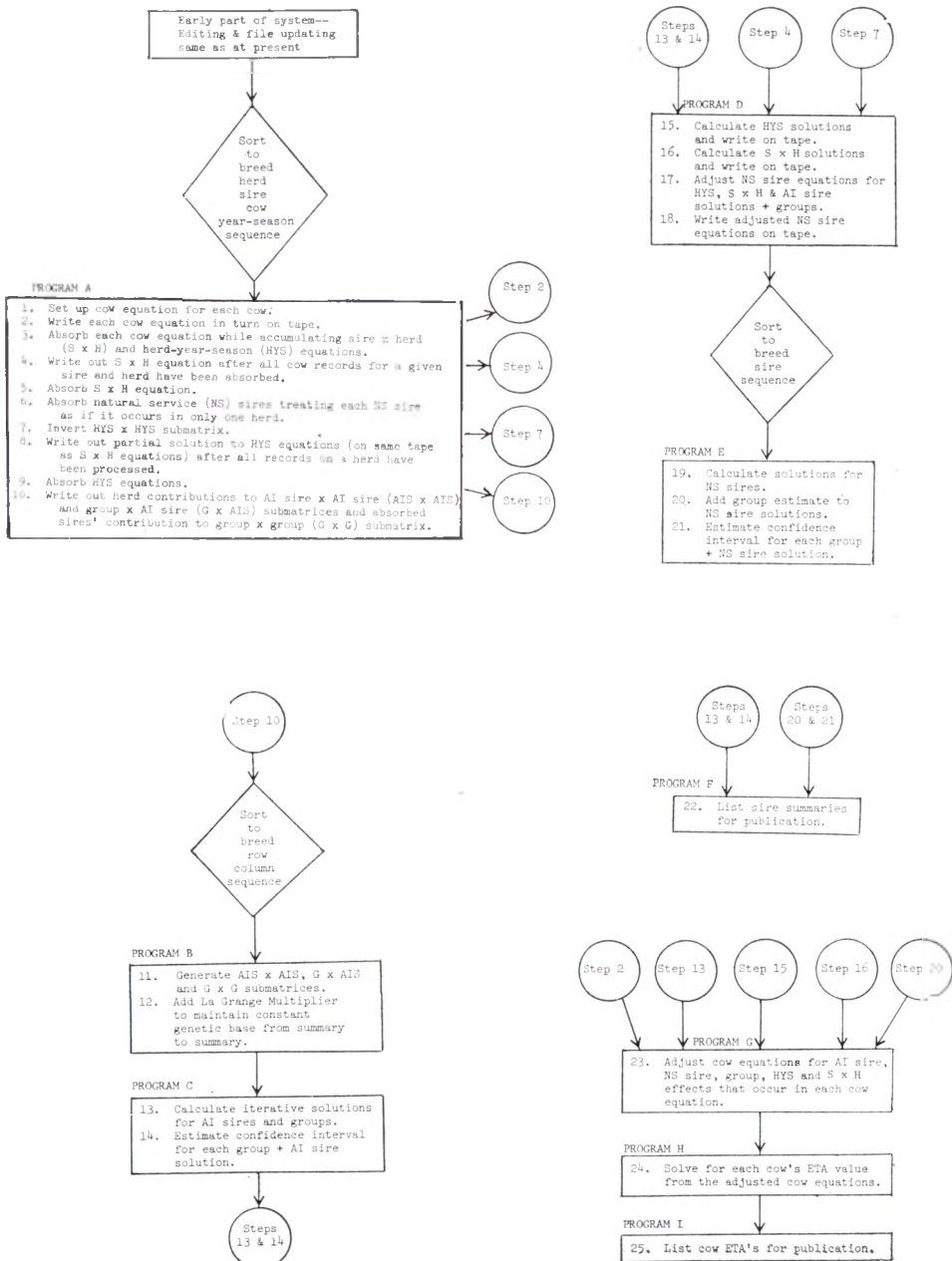


FIG. 2

the *HYS* equations are absorbed (step 9). Finally at the end of each herd two submatrices are written on tape (step 10). These are the *AI* sire \times *AI* sire (*AIS* \times *AIS*) submatrix and the (*G* \times *AIS*) submatrix. A referral to Figure 3 should clarify the procedure followed in steps 1 to 10. During the matrix generation and absorption procedures, the *AIS* contribution to the *G* \times *G* and the *G* \times *AIS* submatrices is ignored in order to save computer time and memory; then this contribution can be calculated later. Therefore, at the end of each herd, these two submatrices contain only the results of the absorption of *NSS*. If there were no *NSS* in the herd (*i. e.*, no sires absorbed), these submatrices contain all zeros.

The first 10 steps shown in Figure 2 are followed, sire by sire within herds and herd by herd, until all input cow records have been processed. The *G* \times *G* submatrix is not written out until after the last herd because it is a relatively small submatrix and can be stored internally.

Group (G)	$\begin{matrix} G \\ \times \\ G \end{matrix}$	$\begin{matrix} G \\ \times \\ AIS \end{matrix}$	$\begin{matrix} G \\ \times \\ HYS \end{matrix}$	$\begin{matrix} G \\ \times \\ NSS \end{matrix}$	$\begin{matrix} G \\ \times \\ S \times H \end{matrix}$	$\begin{matrix} G \\ \times \\ Cow \end{matrix}$
A.I. Sire (AIS)		$\begin{matrix} AIS \\ \times \\ AIS \end{matrix}$	$\begin{matrix} AIS \\ \times \\ HYS \end{matrix}$	$\begin{matrix} AIS \\ \times \\ NSS \end{matrix}$	$\begin{matrix} AIS \\ \times \\ S \times H \end{matrix}$	$\begin{matrix} AIS \\ \times \\ Cow \end{matrix}$
Herd-Year-Season (HYS)			$\begin{matrix} HYS \\ \times \\ HYS \end{matrix}$	$\begin{matrix} HYS \\ \times \\ NSS \end{matrix}$	$\begin{matrix} HYS \\ \times \\ S \times H \end{matrix}$	$\begin{matrix} HYS \\ \times \\ Cow \end{matrix}$
Natural Service Sire (NSS)				$\begin{matrix} NSS \\ \times \\ NSS \end{matrix}$	$\begin{matrix} NSS \\ \times \\ S \times H \end{matrix}$	$\begin{matrix} NSS \\ \times \\ Cow \end{matrix}$
Sire \times Herd (S \times H)					$\begin{matrix} S \times H \\ \times \\ S \times H \end{matrix}$	$\begin{matrix} S \times H \\ \times \\ Cow \end{matrix}$
Cow						$\begin{matrix} Cow \\ \times \\ Cow \end{matrix}$

FIG. 3

The output from program A is then used for a series of sequential operations. The first is shown in programs B and C of Figure 2. This is the solution for the *AI Sire Summaries* and their confidence intervals or Repeatability values.

First, the herd contributions to the *G* \times *AIS* and *AIS* \times *AIS* submatrices are sorted to row and column sequence within breeds. This is shown between steps 10 and 11 of Figure 2. Then the herd contributions to these submatrices are summed to create the *G* \times *AIS* and *AIS* \times *AIS* submatrices and the elements of these submatrices are summed by groups to create the *G* \times *G* submatrix in step 11. This completes the generation of the matrix. This procedure makes it possible to generate a matrix of far greater size than can be stored internally in the computers to which we presently have access. Next, a LA GRANGE Multiplier is used to

maintain a constant genetic base from summary to summary (step 12). The iterative solutions for *AI* sires and for sire groups are calculated in program C (step 13). Following this, the confidence interval or Repeatability value for each *AI* sire's Predicted Difference is calculated (step 14).

The series of operations in programs D and E constitutes the back solution to calculate the *NS* Sire Summaries. The input for this series of operations consists of 3 parts.

1. The *AIS* solutions and confidence intervals or Repeatability values and the group solutions from steps 13 and 14,
2. The $S \times H$ equations from step 4 of program A, and
3. The *HYS* equations from step 7 of program A.

The *HYS* solutions are calculated in program D and written out on tape (step 15). Next, the $S \times H$ solutions are calculated and written out on tape (step 16). Both of these solutions are needed for later use in cow indexing. The *NSS* equations are adjusted for the *HYS*, $S \times H$, and $AIS \times G$ solutions in (step 17). The adjusted *NSS* equations are then written on tape (step 18). These equations are sorted to sire sequence within breed, and the *NSS* solutions are calculated in program E (step 19). The group estimates are added to these solutions (step 20) to give the Predicted Differences for the *NS* sires. Last in this series of operations, the confidence intervals or Repeatability values for the *NS* sire PD's are calculated (step 21).

The final operation in the production of the Sire Summaries program F (step 22), is to combine the *AI* Sire and *NS* Sire Summaries and prepare them for publication.

The last series of operations, at the right of Figure 2, is the calculation of the *Cow Index Values*. There are five inputs for this task. These are the cow equations (step 2), the *AIS* and *G* solutions (step 13), the *HYS* solutions (step 15), the $S \times H$ solutions (step 16), and the *NSS* solutions (step 20). The cow equations are adjusted for each of these effects that occur in each cow equation in program G (step 23). Then the Cow Indexes are calculated from the adjusted cow equations in program H (step 24). The Cow Index Values are then listed for publication in program I (step 25).

As of this writing, we have not determined which of these procedures will be adopted on a long-term basis by *USDA*. This depends on a number of factors such as the dates when they will be ready for implementation, the relative cost in computer time, the hardware and software problems that may be encountered due to the computer environment in which *USDA* operates, and the relative effectiveness of the two methods in the real-world situation. Clearly, an updating of the *USDA* methodology is needed. The genetic progress in dairy cattle populations is probably causing the same difficulties in calculating accurate genetic evaluations in other countries as those in the United States. Therefore, other countries, too, may find these procedures useful for increasing the accuracy of genetic evaluation of dairy cattle.

SUMMARY

Changes in the genetic structure of the dairy cattle population in the United States, caused primarily by genetic improvement, can cause bias in some genetic evaluations calculated by current procedures. Two new methods are described in detail, and flowcharts of computing systems to calculate them are shown. One method is the USDA-DHIA Modified Contemporary Comparison, which uses an iterative procedure to adjust weighted daughter-herdmate deviations for the genetic level of herdmate sires. The iterated deviations are regressed to genetic group means. The second method is the Linear Model Method, which uses very complex statistical and computing procedures to solve extremely large matrices, far beyond direct-solution techniques on any present computer. Both methods are under development by USDA and should eliminate serious sources of bias in the present methodology.

RESUMEN

Los cambios en la estructura genética de las poblaciones de vacuno lechero en Estados Unidos, motivados primordialmente mediante la mejora genética, pueden ser la causa de desviaciones en algunas evaluaciones genéticas calculadas por medio de procedimientos comunes. Se describen detalladamente dos nuevos métodos y se muestran diagramas dinámicos para sistemas de computadores a fin de calcularlos. Uno de los métodos es el USDA-DHIA (Comparación Modificada entre Contemporáneas), que utiliza un procedimiento de corrección para ajustar las desviaciones de las parejas madres/hijas al nivel genético de los reproductores machos. Las diferencias corregidas se comparan con los grupos genéticos medios. El segundo método es el de modelo lineal, que emplea procedimientos estadísticos y de computadores de gran complejidad para resolver matrices extremadamente grandes, y a un nivel muy superior del de las técnicas de solución directa de cualquier computador actual. Ambos métodos están siendo desarrollados por el USDA y deben eliminar fuentes importantes de desviaciones en la metodología actual.

RESUME

Toute modification de la structure génétique du cheptel bovin laitier des Etats-Unis, dont l'origine est surtout une amélioration génétique, risque de fausser les appréciations génétiques calculées par les procédés actuels. Deux nouvelles méthodes sont présentées en détail, accompagnées des fléchages et organigrammes des systèmes d'informatique qui servent à les calculer. La première est la méthode USDA-DHIA (Comparaison Contemporaine Modifiée). Elle recourt à une méthode itérative pour redresser les déviations pondérées mère-fille de l'étable en ce qui concerne le niveau génétique des reproducteurs masculins. La deuxième méthode, dite du Modèle Linéaire, emploie des procédés de calcul informatique et statistique fort complexes pour résoudre de très grandes matrices bien au-delà des techniques dites de solution directe de l'informatique actuelle. L'une et l'autre sont à l'étude au Ministère de l'Agriculture des Etats-Unis; elles devraient permettre de supprimer toutes les sources importantes de déviation dans la méthodologie actuelle.