Improving Accuracy of the United States Genetics Database with a New Editing System for Dairy Records

H. D. NORMAN, L. G. WAITE, G. R. WIGGANS, and L. M. WALTON
Animal Improvement Programs Laboratory
Agricultural Research Service, USDA
Beltsville, MD 20705-2350

ABSTRACT

A new editing system for records used to compute USDA-DHIA genetic evaluations was developed to allow immediate and more complete checking of data. The system uses direct (immediate on-line) access to pedigree and some lactation information to evaluate new data received. Birth dates are checked against parent birth dates and dam calving dates. For most conflicts, existing data are retained, and new data are returned for correction or verification. Records for early lactation can be edited monthly, and data disposition can be determined electronically on-line through telephone access prior to submission. For July 1993 evaluations, nearly 29,000 pedigree records were rejected out of 2.0 million submitted; over 19,000 others were usable after data received were changed to agree with previous information. The most frequent reasons for questionable data in rejected pedigree records were invalid codes or information and conflicts with existing information for animal, sire, and dam identification and animal birth date. Over 158,000 lactation records of 3.2 million submitted with sire identification were rejected; 319,000 others were usable after data were updated to agree with existing information. The four most common reasons for questionable data in rejected lactation records were conflicts for sire or dam identification and animal birth date and invalid sire identification. The new system will decrease turnaround time for correcting unacceptable records.

(**Key words**: dairy, genetics, database, record editing system)

Abbreviation key: AIPL = Animal Improvement Programs Laboratory, **DRPC** = dairy records processing center, **ID** = identification, **RIP** = record in progress.

INTRODUCTION

Genetic improvement is limited by the effective size of the breeding population. Currently, almost all genetic improvement in the US originates from within the country. Foreign genetics that contribute to the US breeding population come primarily from Canada. However, Canadian bulls are sires of only a small percentage of US dairy cattle and an even smaller percentage of the AI bulls sampled in the US. Genetic sources are likely to change as other countries raise their genetic levels and sample more young bulls with AI. Genetics from France and The Netherlands already have begun to be imported into the US.

To enhance genetic improvement, the effective size of the US breeding population can be increased by 1) enrolling more cows in DHI plans, 2) using records from more types of test plans for genetic evaluations, 3) increasing the number of animals with recorded sire and dam identification (**ID**), 4) using unregistered (grade) animals or those from open herdbooks to be dams of young bulls for AI sampling, and 5) incorporating foreign genetics, when superior, into the US breeding population. Increased use of DHI (participation and plans acceptable for genetic evaluation) and improved ID allow more young bulls to be sampled. Improved ID also reduces the cost of sampling a young bull because fewer doses of semen need to be distributed to obtain the desired number or daughters. Permitting grade cows to be bull-dams and inclusion of foreign genetics, when superior, should raise the genetic level of bulls entering AI sampling programs.

Progress is being made in many of these areas because of industry initiative. The num-

Received December 6, 1993.
Accepted May 26, 1994.

ber of cows enrolled in DHI nearly doubled from 1968 to 1984 and has remained steady since (8). The percentage of cows enrolled increased from 16 to 31 for official plans and from 23 to 48 for all plans (7). In January 1993, records from the supervised DHI a.m.-p.m. test plan, code 31 (1), became official (P. Dukas, 1992, personal communication). Although records from this plan have been included for bull evaluations since July 1984 (13), cow evaluations were required to have at least one record from an official test plan. In July 1993, cow evaluations based solely on code 31 records also became eligible for general release. Several other innovative testing plans have been sanctioned by National DHIA (Columbus, OH) for use during a trial period (F. N. Dickinson, 1991, personal communication). Two AI organizations in the US have initiated programs to sample grade bulls (J. R. Thompson and D. Wilson, 1993, personal communications). In January 1993, the Holstein Association of America authorized a second herdbook (2), which could help to bring some of the top grade females into the primary breeding pool of the US. A major effort also is underway to develop and to maintain current formulas for converting genetic evaluations across countries (3, 4, 9, 11) and to identify the top bulls in the leading dairy countries of the world (5, 10).

Missing sire ID is the primary reason that lactation records do not contribute to genetic evaluations. During 1984, 48% of 305-d records from approved test plans were eliminated when genetic evaluations were computed (14); 42% were excluded because of missing sire ID. During January 1992, 35% of 305-d records were eliminated; 31% were excluded because of missing sire ID (USDA, 1993, unpublished results).

US GENETIC DATABASE

Part of the mission of the Animal Improvement Programs Laboratory (**AIPL**), ARS, USDA, is to maintain a highly accurate database of pedigree and lactation information for calculation of USDA-DHIA genetic evaluations for dairy animals. This database provides information for AIPL and university scientists to conduct regional, national, and international genetic research. Accuracy of evaluations and appropriate interpretation of research results are directly dependent on accuracy of the database and the appropriateness of the data editing system used to update it.

Data exchange between AIPL and the dairy industry is shown in Figure 1 as related to the AIPL editing system. Lactation records are provided to AIPL by the dairy records processing centers (**DRPC**), and pedigree information (sire and dam ID and birth and calving dates) are included in these records. Pedigree data also are provided to AIPL by breed associations and by National DHIA with the Verified Identification Program (12), an identity enrollment program primarily for grade cattle. These pedigree files provide an opportunity for additional checking of pedigree accuracy. As a result, detection and correction of some data recording errors are possible within the registered population. Pedigree information often is available for animals enrolled in a herdbook or other ID program up to 2 yr before pedigree data are obtained from lactation records. Identification of parents from breed associations has precedence over that from DRPC. Information from the earliest source has precedence for birth date.

DATA EDITING SYSTEM

A new editing system for constructing the database was implemented for USDA-DHIA genetic evaluations for January 1993. The new system incorporated several editing criteria that AIPL personnel and industry users decided were preferable to previous strategies. In addition, the system was designed for easier use in the field. Direct on-line access to pedigree and some lactation information can be used to evaluate new data. Birth dates are checked against parent birth dates and dam calving dates. For most conflicts, existing data are retained, and new data are returned for correction or verification. Records for early lactation can be edited monthly, and data disposition can be determined electronically prior to submission. The previous editing system was implemented 28 yr ago; therefore, technological advances in data processing since that time could not be used. The previous system was based on the use of sequential data on tape and did not facilitate comprehensive editing.

The new editing system is expected to have its greatest impact on improving the accuracy of animal ID. Recent studies by Meinert and

Norman (7, 8) show that the portion of records from herds enrolled in official test plans that qualified for use in USDA-DHIA genetic evaluations has gradually increased from 44% in 1968 to 64% in 1990. Because most records are excluded because of missing sire ID, some features of the new editing system were designed to encourage the capture of more pedigree information, which in turn would allow an even higher percentage of records to be used for genetic evaluations.

System Features

The new system allows retention of the birth date of each animal and sire and dam ID even if submitted prior to the availability of lactation data. The accuracy of pedigree information is greatest when recorded early in the life of an animal. Pedigree data for females from herds enrolled in management plans (1) are included even though their lactation data are not used for genetic evaluations.

Computer formats are needed to standardize the exchange of information among AIPL, DRPC, breed associations, and the National Association of Animal Breeders. Some simplification was achieved by reducing the number of input formats for individual animal (both pedigree and lactation) data from six for the previous editing system to two for the new system. The format for the pedigree record

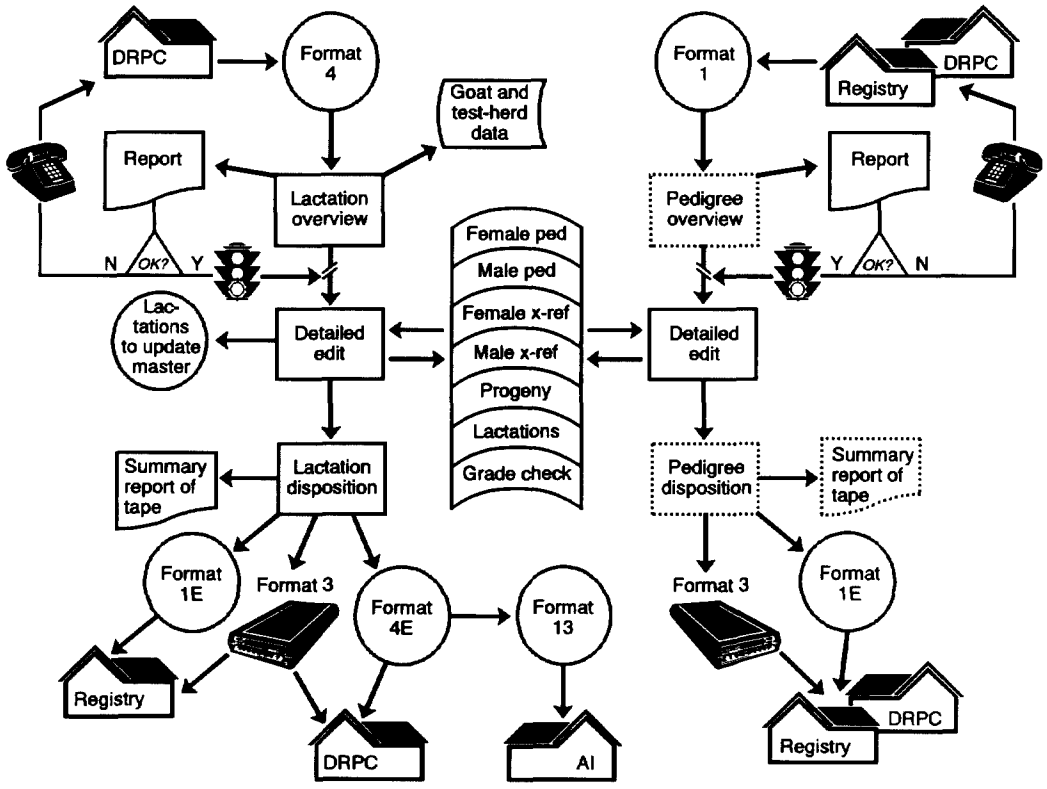

Figure 1. Flow of data exchange between the Animal Improvement Programs Laboratory (AIPL) and dairy industry cooperators as related to the AIPL editing system (circle = data exchange file, rectangle = computer program, curved box = AIPL database, wavy box = report, dashed line = future enhancement, AI = AI organization, DRPC = dairy records processing center, format 1 = pedigree record, format 1E = pedigree error record, format 3 = error reporting record, format 4 = lactation record, format 4E = lactation error record, format 13 = error reporting record for the National Association of Animal Breeders, N = no, ped = pedigree, registry = breed association, x-ref = cross-reference, and Y = yes).

allows the breed associations and DRPC to supply ancestor ID and cross-references to other ID numbers that an animal might have, as well as birth date and information about multiple births and embryo transfer. For grade animals, nearly all pedigree information received by AIPL is provided through their lactation records. Because most US cows are grade animals (6), the lactation record has been the primary source of both pedigree and lactation data.

The cross-referencing feature was developed for animals enrolled in several ID programs. For example, a cow could have ID data from the Holstein-Friesian Association of Canada, the Holstein Association of America, the Red and White Dairy Cattle Association, and the Verified Identification Program, as well as one or more ear tag numbers from the USDA uniform series. Multiple ID for the same animal are detected by checking for other pedigrees with the same sire, dam, and birth date but not coded as twins or embryo transfer. No provisions are available for recording ID of surrogate dams for embryo transfer animals. All information submitted with any of these ID are accessed with a common ID. Preference is given to registration numbers over grade numbers and to US numbers over foreign numbers. For records with valid cross-references, the multiple ID are changed to a single ID in records of the animal and its progeny.

New data are checked immediately on-line (direct access) with previous pedigree and lactation information. If a conflict is detected and the new data have not been submitted with a verification code (verified), the existing data are retained, and the new data are returned for correction or verification. The verification code is used by DRPC to indicate that the record is recognized to have unusual values but thought to be correct. This verification process helps to ensure that conflicting data receive a second examination in the field before existing data are deleted from the database.

Separate edit limits were developed for records that are verified. The range of acceptable values for verified records of production is greater than for unverified records. The objective was to minimize the probability of rejecting legitimate information and to maximize the probability of rejecting data recording errors and inaccurate records. This effort will be successful only through cooperation with the DRPC. If computer programs automatically include verification codes in lactation records without individual examination, coding by AIPL to minimize errors is offset. The verification process can assist producers in obtaining highly accurate information.

The goal of AIPL is to return records with codes to identify questionable data to DRPC and breed associations within 1 wk of receipt. Because of this rapid return of problem records, a DRPC can ensure that dairy producers have an opportunity to correct most records prior to calculation of genetic evaluations. After receipt of a lactation or pedigree record (even for heifers without production data), record disposition codes are generated for DRPC and breed associations if a record contains questionable data. Notification records (code N) indicate that the record was acceptable as submitted but that an unusual condition exists (e.g., no pedigree data supplied by a breed association for a registered animal or parent). These notification records are returned to the appropriate DRPC and breed associations. Changed records (code C) have some data changed or deleted because information from another source (e.g., a cross-reference) is considered to be more reliable; after the change, the record then is accepted. Sire and dam information from breed associations has precedence over information from DRPC for animals recorded by breed associations. For most birth date information, the first information received is retained unless corrected. For production information, the most recent data (i.e., latest processing date) submitted within center have precedence over earlier information. Rejected records (code R) are not accepted because of questionable data accuracy. Code N, C, and R records are returned to DRPC. Code N records may also be sent to a breed association. However, correction of most of these records requires obtaining further information from the producer.

Codes for all questionable data in a record up to six are reported regardless of record disposition. Record disposition is based on these error codes, but the coding assigned to questionable data is not related to that assigned to indicate record disposition. Some questionable data are critical enough to cause immedi-

TABLE 1. Error codes assigned as of May 1994 for questionable data in lactation and pedigree records submitted to the Animal Improvement Programs Laboratory for inclusion in USDA-DHIA genetic evaluations.

Questionable data	Code
Animal identification (ID)	
Invalid breed code or ID	1A
Registered animal with questionable pedigree information	1B
Cross-reference or dual registry available for ID	1C
Grade ID already used by another breed	1D
Animal ID identical to either sire or dam ID number	1F
Grade ID already used by other sex	1G
Invalid cross-reference information based on previous breed code or ID	1H
Invalid name	1J
Invalid unknown grade ID	1K
Invalid deletion or cross-reference request based on pedigree information	1L
Invalid deletion request because of existing lactation information	1M
Sire ID	
Invalid breed code or ID	2A
Registered sire with no pedigree information from breed association	2B
Cross-reference or dual registry available for ID	2C
Grade ID already used by another breed	2D
Sire ID conflicts with existing information	2E
Sire ID identical to dam ID	2F
Grade ID already used by a female	2G
Cross-reference sire ID conflicts with existing sire ID	2H
Dam ID	
Invalid breed code or ID	3A
Registered dam with no pedigree information from breed association	3B
Cross-reference or dual registry available for ID	3C
Grade ID already used by another breed	3D
Dam ID conflicts with existing information	3E
Grade ID already used by a male	3G
Cross-reference dam ID conflicts with existing dam ID	
No resolution through cross-reference in maternal ancestry	3H
Resolution through cross-reference in maternal ancestry	3I
Invalid unknown grade ID	3K
Birth date	
Invalid birth date	4A
Animal birth date conflicts with information for maternal sister	4B
Animal birth date conflicts with dam calving date	4C
Animal birth date conflicts with dam birth date	4D
Animal birth date conflicts with sire birth date	4E
Animal birth date conflicts with existing information	4F
Animal birth date conflicts with first calving date for animal	4G
Cross-reference birth date conflicts with existing birth date	4H
Animal birth date conflicts with progeny birth date	4I
Calving date	
Invalid calving date	5A
Calving age conflicts with birth date	5B
Previous dry days indicated for a first lactation	5C
Excessive number of herd changes for a single calving	5E
Calving date conflicts with existing information	5F
Invalid or unverified deletion request for lactation information	5G
Cross-reference conflict indicated by inability to merge lactation records	5I
Lactation length	
Record in progress with >305 d	6A
Invalid number of days milked three times a day	6B
Invalid DIM	6C
Invalid number of previous days dry	6D
Invalid number of days carried calf	6E
Invalid number of sample days or verified sample days	6F
Invalid processing date	6G

(continued)

TABLE 1. (continued) Error codes assigned as of May 1994 for questionable data in lactation and pedigree records submitted to the Animal Improvement Programs Laboratory for inclusion in USDA-DHIA genetic evaluations.

Questionable data	Code
Production	
Invalid or high last sample day yield	7A
Invalid milk yield	7B
Invalid fat yield	7C
Invalid fat percentage	7D
Invalid protein yield	7E
Invalid protein percentage	7F
Invalid last sample date	7H
Invalid number of somatic cell score (SCS) sample days	7I
Invalid DIM for number of SCS sample days	7J
Invalid SCS	7K
Location and test plan	
Invalid herd code	8A
Invalid test plan code	8B
Invalid processing center code	8C
Other information	
Invalid code for type of lactation record	9A
Invalid weight units code	9B
Invalid species code	9C
Invalid embryo transfer code for lactation record	9D
Invalid termination code for record in progress	9E
Invalid injection code	9F
Invalid lactation status (milking or dry) code	9G
Invalid production verification code	9H
Invalid code for type of pedigree record	9J
Invalid code for pedigree source	9K
Invalid sex code for pedigree record	9L
Invalid species code for pedigree record	9M
Invalid multiple birth code for pedigree record	9N
Invalid version code for pedigree record	9O
Invalid inbreeding percentage for pedigree record	9P
Invalid verification code	9R
Donor dam age >7 yr	9S

ate record rejection; other errors or conflicts that cause a record to be changed may cause rejection only when combined. The seriousness of specific instances of questionable data was decided by AIPL after consultation with dairy industry cooperators. Codes assigned as of May 1994 to indicate questionable data are shown in Table 1. Detailed information on the effect of these error codes on records is available from AIPL.

The new editing system also was designed to facilitate use by cooperators in the field, thus encouraging ID submissions and corrections. Formats are fewer and simpler, and electronic data transfer is possible. Breed associations and DRPC are allowed to check data remotely on-line with AIPL prior to submission to determine the probable disposition of a lactation record. By preparing and transmitting lactation and pedigree records in test mode, a DRPC can receive on-line notification of whether or not a record is acceptable. If a lactation or pedigree record fails to pass AIPL edits, codes that indicate what data were questionable are provided to appropriate industry cooperators. After data correction, the revised record can be tested again on-line for probable disposition. When acceptable, the record can be sent to AIPL for entry into the national database. Use of on-line checking prior to record submission can virtually eliminate the possibility of record rejection during calculation of semiannual genetic evaluations.

Errors in ID are far more likely to cause rejection of lactation records than are production edits. Therefore, most records that would otherwise be rejected can be corrected and retained, which allows the information to be

TABLE 2. Disposition of records with detected errors for 1,969,110 pedigree records with sire identification (ID) that were submitted for calculation of USDA-DHIA genetic evaluations for July 1993.

Record disposition code[1]	Records	Frequency of disposition code among records with detected errors	Frequency of records with detected errors among records with sire ID	Mean detected errors per record
	(no.)	(%)		(no).
N	1377	2.8	.1	1.0
C	19,483	39.1	1.0	1.0
R	28,979	58.1	1.5	1.8
All codes	49,839	100.0	2.5	1.5

[1]N = Notification record (record acceptable as submitted, but herdbook pedigrees missing or abnormal yield values verified), C = changed record (record accepted after some data changed or deleted because information from another source was more reliable), and R = rejected record.

available for genetic evaluations when needed by the producer for management decisions. If a cow was purchased at a high price, her failure to be evaluated hinders effort to market her progeny, particularly sons.

IMPACT ON RECORDS SUBMITTED FOR EVALUATION

Pedigree Information

For USDA-DHIA genetic evaluations for July 1993, 1,969,110 records with sire ID were submitted using the new pedigree format and were processed with the new editing system. Errors were detected in 49,839 records (Table 2): 28,979 records were rejected (code R); 19,483 changed (code C) and 1377 notification (code N) records were included. Questionable data were detected for 2.5% of the records with sire ID, but only 1.5% were excluded. Mean number of detected errors per record was 1.0 for code N and code C records and 1.8 for code R records.

The eight reasons for changing pedigree records with sire ID are in Table 3. Substitution of sire ID because of a cross-reference or dual registry accounted for 70% of all changes; corresponding substitutions for dam ID and cow ID accounted for an additional 23 and 2% of changes, respectively. Conflicts between information submitted for sire ID, dam ID, or birth date and existing information in the database accounted for another 4% of the changes in pedigree records.

The primary reasons for questionable data in rejected pedigree records with sire ID are in Table 4. Invalid sire or dam breed code or ID number and invalid animal birth date were found for 65% of rejected records. Conflicting

TABLE 3. Frequency of reasons for changes in pedigree records with sire identification (ID) that were submitted and then changed for USDA-DHIA genetic evaluations for July 1993.

Reason	Error code	Frequency
		(%)
Cross-reference or dual registry available for sire ID	2C	69.8
Cross-reference or dual registry available for dam ID	3C	23.0
Cross-reference or dual registry available for animal ID	1C	1.9
Dam ID conflicts with existing information	3E	1.7
Sire ID conflicts with existing information	2E	1.3
Animal birth date conflicts with existing information	4F	1.0
Invalid multiple birth code for pedigree record	9N	.7
Invalid verification code	9R	.7

TABLE 4. Ten most frequent reasons for questionable data reported in pedigree records with sire identification (ID) that were submitted for USDA-DHIA genetic evaluations for July 1993 and then rejected.[1]

Reason	Error code	Frequency
		(%)
Invalid sire breed code or ID	2A	31.2
Invalid dam breed code or ID	3A	21.6
Invalid animal birth date	4A	12.1
Animal birth date conflicts with information for maternal sister	4B	10.3
Sire ID conflicts with existing information	2E	7.7
Dam ID conflicts with existing information	3E	5.2
Animal birth date conflicts with existing information	4F	2.4
Animal birth date conflicts with dam calving date	4C	2.0
Invalid multiple birth code for pedigree record	9N	1.4
Invalid animal breed code or ID	1A	1.1

[1]Up to six error codes may be reported for a rejected record; an individual error or conflict may not be the cause of record rejection.

data for sire ID, dam ID, and animal birth date compared with existing information and information submitted for relatives was found for 28% of rejected records. Because up to six error codes may be reported for a rejected record, an individual error or conflict may not have caused the record to be rejected.

Lactation Information

Records for lactations truncated at 305 d or terminated at ≤305 d and records in progress (**RIP**) are provided to AIPL by DRPC as a source of research data, including calculation of genetic evaluations. The 305-d records usually are sent monthly; RIP are sent every month by some DRPC and just prior to semiannual genetic evaluation by others. A requirement by National DHIA that DRPC provide AIPL with individual sample-day data for cows enrolled in innovative test plans (J. Noble, 1991, personal communication) has resulted in more DRPC sending RIP monthly. Previously, AIPL had used only the latest RIP information received from DRPC prior to calculation of genetic evaluations. However, AIPL began monthly editing of RIP data during August 1993 to support early detection of errors.

The impact of the new editing system was examined for the 4,110,898 305-d records and RIP submitted for USDA-DHIA genetic evaluations for July 1993. Missing sire ID was the primary reason that lactation records were not included in calculation of genetic evaluations;

TABLE 5. Disposition of records with detected errors for 3,228,540 lactation records with sire identification (ID) that were submitted for USDA-DHIA genetic evaluations for July 1993.

Record disposition code[1]	Records	Frequency of disposition code among records with detected errors	Frequency of records with detected errors among records with sire ID	Mean detected errors per record
	(no.)	(%)	(%)	(no.)
N	2493	.5	.1	1.1
C	319,063	66.5	9.9	1.3
R	158,446	33.0	4.9	2.2
All codes	480,022	100.0	14.9	1.6

[1]N = Notification record (record acceptable as submitted, but herdbook pedigrees missing or abnormal yield values verified), C = changed record (record accepted after some data changed or deleted because information from another source was more reliable), and R = rejected record.

TABLE 6. Ten most frequent reasons for changes in lactation records with sire identification (ID) that were submitted for USDA-DHIA genetic evaluations for July 1993.

Reason	Error code	Frequency
		(%)
Invalid processing date	6G	33.5
Invalid number of sample days or verified sample days	6F	24.6
Invalid lactation status (milking or dry) code	9G	17.9
Dam ID conflicts with existing information	3E	6.4
Animal birth date conflicts with existing information	4F	5.3
Sire ID conflicts with existing information	2E	3.5
Cross-reference or dual registry available for sire ID	2C	2.6
Cross-reference or dual registry available for animal ID	1C	2.4
Cross-reference or dual registry available for dam ID	3C	1.8
Invalid or high last sample day yield	7A	.9

21% were excluded because of no sire ID in the incoming record and no sire ID previously reported by a breed association or DRPC. At least two DRPC do not send RIP without sire ID to AIPL. Therefore, the actual percentage of DHI records excluded because of missing sire ID is higher than 21%, probably about 26 or 27%. Although AIPL historically had excluded incoming records with missing sire ID, the new editing system allowed recovery of a small percentage of these records because of sire ID reported previously. For July 1993 evaluations, 1367 animals had a sire ID of zero changed to a previously reported ID. These included 788 registered cows, 7 Verified Identification Program cows, and 595 cows in the Identified Holstein Females program of the Holstein Association of America. Records with unrecoverable sire ID are not returned to DRPC because of the many records with little opportunity for correction and because DRPC already are aware that sire information is missing for these records. An additional 1141 animals had a dam ID of zero changed to an ID previously reported by a breed association. All these animals were registered cows. If DRPC could provide pedigree information for grade animals early in life, the number of records that could be recovered using previously known sire and dam information would be expanded greatly. For any record submitted without either sire or dam ID, the DRPC are notified of probable sire or dam ID if AIPL has previous information for that animal.

TABLE 7. Ten most frequent reasons for questionable data reported in lactation records with sire identification (ID) that were submitted for USDA-DHIA genetic evaluations for July 1993 and then rejected.[1]

Reason	Error code	Frequency
		(%)
Sire ID conflicts with existing information	2E	14.7
Animal birth date conflicts with existing information	4F	14.1
Dam ID conflicts with existing information	3E	13.0
Invalid sire breed code or ID	2A	9.7
Animal birth date conflicts with information for maternal sister	4B	6.6
Calving date conflicts with existing information	5F	6.5
Invalid number of sample days or verified sample days	6F	5.8
Invalid lactation status (milking or dry) code	9G	4.0
Invalid fat yield	7C	3.8
Invalid dam breed code or ID	3A	3.6

[1]Up to six error codes may be reported for a rejected record; an individual error or conflict may not be the cause of record rejection.

The 3.2 million lactation records with sire ID were processed with the complete editing system to determine how many records were usable for genetic evaluation. Errors were detected in 480,022 records (Table 5): 158,466 records were rejected (code R); 319,063 changed (code C) and 2493 notification (code N) records were used to calculate genetic evaluations for July 1993. Although questionable data were detected in 14.9% of records with sire ID, only 4.9% were excluded as unacceptable. The mean number of detected errors per record was 1.1 for code N records, 1.3 for code C records, and 2.2 for code R records.

The 10 primary reasons for changing lactation records with sire ID are in Table 6. Technical concerns related to processing those records accounted for the majority of changes. The 3 most frequent reasons for changing records accounted for 76% of changes and were related to computer processing issues. An error in recording the processing date for all records from one DRPC resulted in a large number of changed records. For this situation, the processing date was changed to make it compatible with calving date and lactation length. The second most frequent reason for changing records was an invalid number of sample days or verified sample days. For this situation, the code for number of sample days is removed and left blank. If an invalid code for milking status was included in the record, it was changed to 0 to indicate that the code would not be used. Of the remaining 7 frequent reasons for changing records, 6 concerned conflicts between input data and existing information (including cross-references) and accounted for 22% of changes.

The 10 most common reasons for questionable data in rejected lactation records with sire ID are in Table 7. Conflicts of sire, dam, and animal information with existing information or information submitted for relatives were found for 55% of the rejected records. Invalid coding or invalid information was found for 27% of rejections.

DISCUSSION

Some interest has been expressed by the dairy industry in receiving notification of ID errors and conflicts in lactation records on the same day that an animal enters the milking herd. This rapid checking may be possible because of advances in computer technology. The best time to detect ID errors is at the earliest possible opportunity; i.e., when the dam is bred, when the animal is enrolled in a calf program, or when the cow enters the milking herd.

A modest improvement in ancestor ID may be possible, especially for grade animals, if DRPC provide information earlier in an animal's life. Although many DRPC provide dairy producers with the option of recording breeding information and reporting heifer inventory, the extent to which this information is transferred automatically to the milking herd has not been documented. A survey on the completeness and flexibility of DRPC programs and the extent of their use would be beneficial for determining how information from these programs could be integrated into the new editing system to increase record usability. Increased participation in the DRPC programs to record breeding and heifer information could enhance genetic improvement, and AI organizations could develop incentive programs for progeny test herds that enroll in such programs.

The impact of using records without sire ID for genetic evaluation needs further investigation. Although their inclusion could increase the number of contemporaries, the effect of potential biases on overall accuracy of evaluations is unknown.

With the implementation of the new AIPL editing system and the current interest of AI organizations in improving ID, now may be the appropriate time to emphasize improved ID accuracy, thereby increasing the effective size of the US breeding population and increasing the rate of genetic improvement. Adoption of a single ID system industrywide could decrease the number of records rejected because of conflicts among multiple ID systems. The AIPL editing system will continue to be enhanced as improvements are recognized.

ACKNOWLEDGMENTS

The cooperation of the US dairy industry in supplying production and pedigree data through the National Cooperative DHI Program is appreciated greatly. Appreciation is extended to K. R. Butcher, J. S. Clay, and S. M. Hubbard, as well as to the *Journal of Dairy*

Science reviewers for suggestions to improve the manuscript.

REFERENCES

1 Benson, R. H. 1985. The NCDHIP record plans. Natl. Coop. DHI Prog. Handbook, Fact Sheet A-4. Ext. Serv., USDA, Washington, DC.

2 Holstein Association of America. 1993. Qualified Herdbook Rules of Holstein Association of America as Accepted April 1, 1993. Holstein Assoc. Am., Brattleboro, VT.

3 International Bull Evaluation Service. 1990. Recommended procedures for international use of sire proofs. Int. Bull Eval. Serv. Bull. No. 4. Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.

4 International Bull Evaluation Service. 1992. Sire evaluation procedures for dairy production traits practiced in various countries 1992. Int. Bull Eval. Serv. Bull. No. 5. Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.

5 Koerhuis, A., F. van der Schans, R. L. Powell, and G. R. Wiggans. 1991. How good are the U.S. Holstein young sires sampled in other countries? Holstein World 88:2015.

6 Majeskie, J. L. 1992. Status of United States dairy cattle. Natl. Coop. DHI Prog. Handbook, Fact Sheet K-7. Ext. Serv., USDA, Washington, DC.

7 Meinert, T. R., and H. D. Norman. 1992. Participation in the National Cooperative Dairy Herd Improvement Program and usability of records in USDA-DHIA genetic evaluations. J. Dairy Sci. 75(Suppl. 1): 244.(Abstr.)

8 Meinert, T. R., and H. D. Norman. 1994. Records contributing to national genetic evaluations by year, registry status, herd size, and region. J. Dairy Sci. 77: 1466.

9 Philipsson, J. 1987. Standards and procedures for international genetic evaluations of dairy cattle. J. Dairy Sci. 70:418.

10 Powell, R. L., and M. Sieber. 1991. Which country has the best Holstein bulls in 1991? Holstein World 88:576.

11 Powell, R. L., and M. Sieber. 1992. Direct and indirect conversion of bull evaluations for yield traits between countries. J. Dairy Sci. 75:1138.

12 Voelker, D. E. 1981. Dairy Herd Improvement Associations. J. Dairy Sci. 64:875.

13 Wiggans, G. R., and S. M. Hubbard. 1986. Changes in USDA-DHIA genetic evaluation procedures since January 1984. Natl. Coop. DHI Prog. Handbook, Fact Sheet H-13. Ext. Serv., USDA, Washington, DC.

14 Wiggans, G. R., and D. R. Ostrander. 1985. Counts of 305-day lactation records eliminated from genetic evaluations for 1984. Natl. Coop. DHI Prog. Handbook, Fact Sheet H-10. Ext. Serv., USDA, Washington, DC.