

A Computationally Feasible Test Day Model for Genetic Evaluation of Yield Traits in the United States¹

G. R. WIGGANS* and M. E. GODDARD†

*Animal Improvement Programs Laboratory,
Agricultural Research Service,
USDA, Beltsville, MD 20705-2350

†University of New England,
Armidale, New South Wales, Australia

ABSTRACT

A multitrait analysis of test day yields is proposed that includes 60 traits [3 yield traits (milk, fat, and protein), 2 parity groups (first and later) per yield trait, and 10 stages of lactation per parity]. To reduce the computations needed for the 60 traits, test day effects are estimated within the herd before analysis across herds, the rank of the genetic (co)variance matrix is reduced, a canonical transformation is used with missing values replaced by their expectations, and a repeatability model is applied to allow inclusion of parities after second. Historical 305-d records are included through their correlations with test day effects. Possible benefits from this model include 1) more accurate estimation of environmental effects from including the influence of particular days of recording, 2) optimal use of information from all test days (especially for lactations with long intervals from calving to first test or between tests), 3) improved accuracy of evaluations for component yields through contributions from information for milk yield, and 4) greater stability of bull evaluations from accounting for genetic differences among daughters in the shape of lactation curve and maturity rate.

(**Key words:** test day model, genetic evaluation, multitrait analysis, yield traits)

Abbreviation key: TD = test day.

INTRODUCTION

Individual yields on test day (TD) can be the basis of a genetic evaluation system instead of estimates of yield during the first 305 d of lactation. Lactation yields typically are estimated from monthly measurements of milk volumes and analysis of milk samples for fat and protein percentages (24). Recent work (15, 21, 23) with TD data has confirmed the greater

precision that can be achieved in accounting for environmental effects by including a TD effect. Van Tassell et al. (23) found heritability increases of 12% for milk yield, 11% for fat yield, and 17% for protein yield; correlations among traits increased 2 to 14%. At 20, 150, and 280 DIM, Ptak and Schaeffer (15) found a reduction in residual variances for milk yield from 11.7, 4.5, and 7.6 kg² to 8.4, 3.3, and 5.1 kg², respectively, for a model that included effect of herd and TD rather than effect of herd, year, and season.

A recent review (22) reports two approaches for use of TD data: 1) correction for environmental influences at the TD level followed by aggregation into a lactation measure for later processing and 2) fitting a model to TD data directly. Both approaches provide a more accurate accounting of environmental effects. Lactation records of contemporaries often do not include all of the same TD because of differences in calving date and lactation length, but the TD effect in a TD model is an estimate of the environmental effect of the same day for all cows that contribute to the effect. Including a TD effect increases the precision of accounting for environmental effects. Direct analysis of TD yields is computationally more demanding but allows for genetic differences in lactation curves.

The cost of milk recording is being reduced by making fewer measurements, which causes longer intervals between milk weighings and less frequent collection of milk samples. Regardless of the length of the interval between tests, a TD model can appropriately weight the recorded TD information by considering the covariances among TD yields. Yields from two TD that are close and, therefore, highly correlated would not contribute as much information as yields from two more distant TD. Use of TD data would allow the use of information from lactations with long intervals between milk recordings because estimation of yields for unrecorded intervals would not be required. Although a TD model cannot overcome the loss in accuracy from fewer TD, it does allow yields from any combination of TD to be included appropriately.

Received July 15, 1996.

Accepted February 10, 1997.

¹Invited paper.

Several countries currently have evaluation systems that use TD data that have been adjusted and then combined into a lactation measure. The Australian system, implemented in 1984, estimates and removes TD effects before creating a lactation measure (10). In the northeastern US, effects of TD, including age and lactation stage, are estimated within a herd to create adjusted lactation measures for an analysis across herds (1, 3). A method recently developed in New Zealand (9) combines TD into a 270-d lactation yield that weights the individual TD according to the correlations among them. The correlation declines linearly as days between TD increase.

In Canada, TD yields are analyzed directly for genetic evaluation of some traits. Yield traits of dairy goats (19) and SCS of dairy cattle (16) are analyzed with a model that estimates regression coefficients for lactation curves within 24 groups by parity, age at parturition, and season of parturition for the population being evaluated. The Canadian Genetic Evaluation Board has recommended implementation in 1998 (L. R. Schaeffer, 1996, personal communication) of a TD model for yield traits of dairy cattle. The multitrait model would include milk, fat, and protein yields and SCS for each of the first three parities for a total of 12 traits. Coefficients of the lactation curve would be fit for each animal as random effects (7, 8) using the additive relationships among animals and the genetic covariances among the random regression coefficients. The presence of only one trait would be sufficient for a record to be used. This approach would allow estimation of genetic differences in persistency for yield traits of dairy cattle.

Interest is increasing worldwide in TD models. Genetic parameters of TD yield have been estimated for Finnish Ayrshires (14) and Spanish Holstein-Friesians (18). A TD model for multiple parities has been applied to predict genetic merit for yield traits of German Holsteins (17).

A multitrait analysis allows information for one trait to contribute to the accuracy of correlated traits and also allows a correlation of <1 to be assumed between observations that have been treated as the same trait. A multitrait analysis that includes milk, fat, and protein yields allows measurements of milk volume to contribute to the accuracy of evaluations of milk components because of the correlations between components and milk volume. This characteristic is particularly useful when component determination is less frequent than measurement of milk volume. More frequent collection of milk volumes and less frequent determination of milk components are expected as the use of inline equipment for recording daily milk volumes and the expense and incon-

venience of component sampling increase. In this environment, a multitrait system can provide more information on genetic merit for component yields than a single-trait system can.

A repeatability model is typically used for genetic evaluation of dairy cattle. Such a model assumes that each lactation is an expression of the same genetic trait. Estimated genetic correlations between first and second parities ranged from 0.88 to >1.00 for milk, fat, and protein of Australian Black and White cattle (12); however, genetic correlations between parities of Spanish Holsteins ranged from 0.65 to 0.91 among parities 1 through 4 for milk, fat, and protein (4). For TD milk yields, genetic correlations between parities ranged from 0.52 to 0.88 among parities 1 through 3 (18). Those correlations of <1 indicate some divergence from the assumption that milk yield is the same trait across parity. A multitrait model that includes parities as separate traits can accommodate this correlation structure. The benefit of a multitrait analysis over several single-trait analyses is greatest when some traits have missing observations or when heritability of all traits is not the same and thus genetic and phenotypic correlations are not equal. Similarly, when separate genetic effects are fit to lactation stages or when coefficients of the lactation curve for each cow are fit by random regression (6), genetic differences in the shape of the lactation curve (persistency) can be detected (13, 18, 21, 22).

A plan follows for a multitrait TD model for calculation of genetic evaluations for yield traits of US dairy cattle that is patterned on the method currently being developed for Australia (25). This approach includes a full multitrait model with a more realistic (co)variance structure than the current model for USDA-DHIA genetic evaluations and is computationally feasible to implement.

MATERIALS AND METHODS

Data

Nearly complete TD data for calvings since January 1, 1990 have been extracted from records provided routinely for USDA-DHIA genetic evaluations and from archive files provided by dairy records processing centers and universities. The TD information is stored in a database table that includes cow identification, calving date, DIM, milking frequency, milk yield, fat and protein percentages, and SCS. Herd code is stored in a lactation table, and information on the type of milk recording is stored in a table for each herd test date. Because cows sometimes change herds and because most of the archive data are stored based

on lactation or cow, an incorrect herd code sometimes is associated with a TD record. The data table based on herd and TD includes the number of cows with yield on that TD, which allows herd code errors to be detected. Lactation yields (305-d records) can be used for calvings before 1990 and for herds with limited TD data.

Statistical Model

A proposed model for TD yield is

$$y_{ijklmnopqrstu} = htp_{ijkp} + a_{ijmpq} + dim_{ijmps} \\ + pdo_{ijmr} + do_{ijmu} + hys_{ijlmpt} \\ + hs_{ijopt} + bv_{inpt} + pe_{inpt} \\ + e_{ijklmnopqrstu}$$

where $y_{ijklmnopqrstu}$ is TD yield of trait i (milk, fat, or protein) in herd j on TD k in season l of year m from cow n that was sired by bull o and that calved for parity p at age q after r days open and currently has s DIM (t mo in milk) and u days open; htp is fixed effect of herd, TD, and parity; a is fixed effect of age at calving within parity; dim is fixed effect of DIM within parity; pdo is fixed effect of previous days open; do is fixed effect of current days open; hys is fixed effect of herd-year-season; hs is random effect of interaction of herd and sire; bv is random breeding value of the cow; pe is random effect of permanent environment; and e is random residual pertaining to TD yield.

The model assumes distinct traits by months in milk and does not allow for two tests during the same month. Months in milk can be determined by assigning TD to 1 of 10 30-d groups starting at 5 DIM. To avoid loss of data when >1 TD would be assigned to the same group, one of the TD can be assigned to an adjacent group if the group is empty and the TD is within 10 d of that group; otherwise, yields can be averaged. An alternative is to reduce the group size to 20 d so that multiple observations per group would be rare. The htp effect is across months in milk to ensure sufficient observations per TD, but the hys effect is specific for each trait during the analysis across herds to allow for differences in the effect of season on the traits for individual months in milk. Between two and four seasons are proposed. Additional research will determine the best representations of dim , pdo , and do effects; individual effects by herd and parity grouping may be appropriate. For hys , hs , bv , and pe , parities are grouped as first or later parity.

This multitrait model includes 60 traits: 3 yield traits (milk, fat, and protein), two parity groups (first and later) per yield trait, and 10 stages of

lactation per parity. The TD yields within a lactation are considered to be separate traits. Data from parities 3 through 5 are considered to be repeated observations of the 30 traits for parity 2 (later lactations). This repeatability model allows for multiple parities; each lactation conceptually contains all 60 traits but with missing observations for ≥ 30 traits.

The variance components that are required for the 60 traits would be time-consuming to estimate. Some structure could be imposed by forcing regular decay in the correlation between TD as the interval between them increases. The effect of interaction of herd and sire is included to limit the evaluations of bulls with daughters in only a few herds. To achieve this dampening, the variance assigned for the hs effect will be larger than the hs variance estimated from the data. An arbitrary portion of the pe variance can be assigned to hs variance. The (co)variance matrices must include 305-d yield so that covariance with each month in milk is available.

Computational Method

First, data are adjusted (adj) using solutions estimated for hys , hs , bv , and pe effects from a previous ($prev$) analysis across herds:

$$y_{adj_{prev}} = y - hys_{prev} - hs_{prev} - bv_{prev} - pe_{prev}.$$

Yield from a 305-d record would be adjusted similarly to TD yield. Then, effects for htp , a , dim , pdo , and do are estimated within the herd (wh):

$$y_{adj_{prev}} = htp_{wh} + a_{wh} + dim_{wh} + pdo_{wh} + do_{wh} + e_{wh}.$$

Effects of a_{wh} , dim_{wh} , and do_{wh} are estimated with an across-herd component to incorporate prior information (9). Data then are adjusted for effects within the herd:

$$y_{adj_{wh}} = y - htp_{wh} - a_{wh} - dim_{wh} - pdo_{wh} - do_{wh},$$

followed by estimation of hys , hs , bv , and pe effects across herds with a multitrait animal model. Six linear functions are imposed on the genetic (co)variance matrix to reduce its rank, and a canonical transformation is applied to create the known contributions to 6 traits with genetic variance >0 . During each round of iteration, the contribution to those variables from the missing observations is calculated from the solutions from the previous round of iteration. After convergence, solutions for the 6 traits are backtransformed to the original 60 traits and

allowed to affect solutions for htp, a, dim, pdo, and do effects within the herd until convergence is reached. The final step is to calculate reliabilities.

Reduction of the rank of the genetic (co)variance matrix by imposing six linear functions.

Selection decisions for dairy cattle do not require specific estimates for genetic effects of monthly TD for all traits. Functions of those TD effects often are sufficient. However, by defining 60 traits, the covariance structure can be accommodated.

The number of computations that are needed for prediction of EBV can be reduced if fewer traits are analyzed. The rank of the genetic (co)variance matrix (\mathbf{G}), which has order 60, can be reduced to 6 so that \mathbf{G} retains only information for lactation yield, aspects of the lactation curve, and yield differences due to parity. Because of the rank reduction, only 6 canonical traits need to be analyzed after canonical transformation.

The goal is for the matrix with reduced rank (\mathbf{G}^*) to include nearly all of the variation in lactation yields of milk, fat, and protein. The largest eigenvectors of the canonical decomposition of \mathbf{G} and the phenotypic (co)variance matrix are not proposed because selection response with variance components from Australian data that used the first five eigenvectors led to some loss of accuracy in the EBV for lactation yield (K. Meyer, 1995, personal communication).

Therefore, \mathbf{G}^* was defined by six linear combinations of TD traits that are of interest in the selection objective: 1) Σbv_{milk} across parities and TD, 2) Σbv_{fat} across parities and TD, 3) $\Sigma bv_{\text{protein}}$ across parities and TD, 4) linear trend in bv_{milk} , bv_{fat} , and bv_{protein} across TD for both parity groups, 5) quadratic trend in bv_{milk} , bv_{fat} , and bv_{protein} across TD for both parity groups, and 6) (bv_{milk} , bv_{fat} , and bv_{protein} for first parity) - (bv_{milk} , bv_{fat} , and bv_{protein} for later parity). The first three combinations provide information on lactation yield for milk, fat, and protein. The linear and quadratic weightings of yield traits are expressed in units of standard deviation and provide information on persistency. The difference between first and later parities also is expressed in standard deviation units and is an indicator of maturity rate.

Canonical transformation. If no data are missing, a canonical transformation can be used to convert a multitrait analysis into single-trait analyses, which substantially reduces computational requirements. Each lactation can contribute a maximum of 30 TD yields for first or later parity; the other 30 traits, and perhaps more, are missing. Lactations with <10 TD would have more missing values. For a lactation that is represented by only a 305-d record, all 60 traits

would be missing, but the 6 canonical traits would contain known data from the 305-d record derived from its correlations with the 60 TD traits. Ducrocq and Besbes (2) extended the application of canonical transformation to situations in which not all traits were measured by replacing missing values with their expectations during each round of iteration. Returning to the original scale during iteration is not necessary.

A canonical diagonalization depends on simultaneously diagonalizing the (co)variance matrices for random effects. Although the genetic and residual (co)variance matrices can be diagonalized, a repeatability model also includes permanent environment as a random effect. Lin and Smith (11) describe a method of approximate diagonalization that allows the canonical transformation to be extended to models with multiple random effects. The success of this approach is measured by the relative size of the off-diagonal elements after diagonalization. If the (co)variance matrix of hs is defined as proportional to the (co)variance matrix of pe, the matrix will diagonalize as effectively as (co)variance of pe.

The residual covariance matrix and \mathbf{G}^* are used in creating a transformation matrix to construct a canonical transformation of the 60 traits. This procedure generates 60 transformed traits that are environmentally and genetically uncorrelated. However, because the rank of \mathbf{G}^* is 6, only six canonical transformations have any genetic variance. Therefore, the 60-trait analysis is replaced by six single-trait analyses on the transformed scale.

Gauss-Seidel iteration. Iteration consumes most of the time that is required to calculate evaluations. Gauss-Seidel iteration may converge more rapidly than second-order Jacobi iteration but does require that solutions from the appropriate iteration be used when accumulating the right-hand sides. A linked list can be used to collect the identification of mates as described by Strandén and Mäntysaari (20). The table that stores pedigree data is filled so that the parent with the earlier birth date is in the first column. By processing animal solutions in order of animal number, contributions from mates with lower numbers can be added to the accumulator immediately before the new solution is calculated by using solutions from the current round instead of the previous round as happens in second-order Jacobi iteration. This approach makes the use of Gauss-Seidel iteration nearly as simple as the use of second-order Jacobi iteration and may require fewer rounds of iteration.

Use of 305-d records. To avoid selection bias and to improve estimates of genetic trend, 305-d records

for calvings before 1990 can be included. Those records can contribute to estimation of missing TD yields through correlations of 305-d yield with TD effects. For calvings after 1990, TD data are incomplete for some herds; for herd TD for which >25% of TD yields are missing, available TD information can be supplemented with 305-d yields. A 305-d yield can be expressed on a daily yield basis so that the contribution from a 305-d record is similar in magnitude to TD yield. The 305-d yields may need to be adjusted to the same standard DIM as the TD yields to make them comparable.

Reliability. The pattern of missing values and correlations between traits should be reflected in the reliabilities. For example, four tests in consecutive stages provide less information than four tests that are evenly spaced across lactation. Gengler and Misztal (5) have proposed a method that has acceptable accuracy if the proportion of missing records or the correlation between traits is moderate. However, their procedure may not be appropriate for this 60-trait design, which has over half of the traits missing. Approximations that are suitable for this data structure are needed, and methods for estimating reliability for multitrait analysis may be adaptable (6).

DISCUSSION

The proposed evaluation method based on TD data has several beneficial features. A canonical transformation is used despite unequal design matrices through a two-step solution strategy. This feature is advantageous for genetic evaluation of dairy cattle because the first analysis step can be computed within the herd, which reduces the volume of data at the second step (across herds) to a level comparable with traditional analyses for lactation yields. Another useful feature is that the htp effect includes all lactation stages for a given TD and parity, which allows more observations to contribute to estimation of this effect; however, using the same htp definition does require that adjustment for DIM be appropriate across lactation stages. Finally, when many traits are closely related, use of a genetic (co)variance matrix of reduced rank saves computational time.

Solutions from the TD model should be included in an index to make maximum genetic progress. Solutions for the two parity groups could be weighted by the group frequencies in the population. Similarly, evaluations by month in milk could be weighted equally, or later months could be weighted more heavily to select for persistency.

If genetic differences exist in the shape of lactation curves or in maturity rate, this multitrait evaluation, which includes separate traits for lactation stage and parity, might result in more stable evaluations. Correlations between traits are <1, which restricts the degree that one stage or parity can influence the genetic values for another and reduces the impact of large amounts of data that are concentrated in early stages or early parities. For example, if second crop daughters of a bull tended to have high peak yields during early lactation, their projected 305-d yields would overestimate actual yield, and the bull evaluation would be too high. With a multitrait model for TD, the information from later stages of first crop daughters would not be overwhelmed by the many observations for the early stages of second crop daughters; therefore, the evaluation would be more stable.

CONCLUSIONS

A TD model is possible for the US dairy industry. Evaluations from such a model are expected to be more accurate because of better accounting for environmental effects. Those evaluations also should be more stable because of accounting for genetic differences in maturity rate and persistency. The proposed approach for implementation includes 305-d yields (in addition to TD yields) to account appropriately for genetic trend. A TD model is best able to accommodate the wide variation in milk recording that is developing in an effort to reduce costs. All correlations among milk, fat, and protein yields can be used. Any number and distribution of TD within parity as well as different frequencies of recording volume and components can be accommodated. Correlations of <1 among TD and between parities allows appropriate weighting of information.

REFERENCES

- 1 Cooperative Extension. 1995. The Sire Report, July 1995, Northeast A.I. Multiple Trait Cow & Sire Evaluation. Cornell Coop. Ext., Dep. Anim. Sci., Cornell Univ., Ithaca, NY.
- 2 Ducrocq, V., and B. Besbes. 1993. Solution of multiple trait animal models with missing data on some traits. *J. Anim. Breed. Genet.* 110:81.
- 3 Everett, R. W., F. Schmidt, and L. H. Wadell. 1994. A test-day model for monitoring management and genetics in dairy cattle. *J. Dairy Sci.* 77(Suppl. 1):267.(Abstr.)
- 4 García-Cortés, L. A., C. Moreno, L. Varona, M. Rico, and J. Altarriba. 1995. (Co)variance component estimation of yield traits between different lactations using an animal model. *Livest. Prod. Sci.* 43:111.
- 5 Gengler, N., and I. Misztal. 1996. Approximation of reliability for multiple-trait animal models with missing data by canonical transformation. *J. Dairy Sci.* 79:317.
- 6 Graser, H.-U., and B. Tier. 1997. Applying the concept of number of effective progeny to approximate accuracies of predictions

- derived from multiple trait analyses. *Proc. Assoc. Adv. Anim. Breeding Genet.*, Dubbo, New South Wales, Australia, 12:547.
- 7 Jamrozik, J., L. R. Schaeffer, and J.C.M. Dekkers. 1996. Random regression models for production traits in Canadian Holsteins. *Proc. Open Session INTERBULL Annu. Mtg.*, Veldhoven, The Netherlands, June 23–24, 1996. *Int. Bull. Eval. Serv. Bull. No. 14:124.* Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.
 - 8 Jamrozik, J., L. R. Schaeffer, and J.C.M. Dekkers. 1997. Genetic evaluation of dairy cattle using test day yields and random regression model. *J. Dairy Sci.* 80:1217.
 - 9 Johnson, D. L. 1996. Estimation of lactation yield from repeated measures of test day yields. *Proc. N.Z. Soc. Anim. Prod.* 56:16.
 - 10 Jones, L. P., and M. E. Goddard. 1990. Five years experience with the animal model for dairy evaluations in Australia. *Proc. 4th World Congr. Genet. Appl. Livest. Prod.*, Edinburgh, Scotland, XIII:382.
 - 11 Lin, C. Y., and S. P. Smith. 1990. Transformation of multitrait to unitrait mixed model analysis of data with multiple random effects. *J. Dairy Sci.* 73:2494.
 - 12 Meyer, K. 1985. Genetic parameters for dairy production of Australian Black and White cows. *Livest. Prod. Sci.* 12:205.
 - 13 Meyer, K., H.-U. Graser, and K. Hammond. 1989. Estimates of genetic parameters for first lactation test day production of Australian Black and White cows. *Livest. Prod. Sci.* 21:177.
 - 14 Pösö, J., E. A. Mäntysaari, and A. Kettunen. 1996. Estimation of genetic parameters of test day production in Finnish Ayrshire cows. *Proc. Open Session INTERBULL Annu. Mtg.*, Veldhoven, The Netherlands, June 23–24, 1996. *Int. Bull. Eval. Serv. Bull. No. 14:45.* Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.
 - 15 Ptak, E., and L. R. Schaeffer. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci.* 34: 23.
 - 16 Reents, R., J.C.M. Dekkers, and L. R. Schaeffer. 1995. Genetic evaluation for somatic cell score with a test day model for multiple lactations. *J. Dairy Sci.* 78:2858.
 - 17 Reents, R., and L. Dopp. 1996. Genetic evaluation for dairy production traits with a test day model for multiple lactations. *Proc. Open Session INTERBULL Annu. Mtg.*, Veldhoven, The Netherlands, June 23–24, 1996. *Int. Bull. Eval. Serv. Bull. No. 14:113.* Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.
 - 18 Rekaya, R., R. Béjar, M. J. Carabaño, and R. Alenda. 1995. Genetic parameters for test day measurements in Spanish Holstein-Friesian. *Proc. Open Session INTERBULL Annu. Mtg.*, Prague, Czech Republic, Sept. 7–8, 1995. *Int. Bull. Eval. Serv. Bull. No. 11.* Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.
 - 19 Schaeffer, L. R., and B. P. Sullivan. 1994. Genetic evaluation of dairy goats using test day yields. *Proc. 5th World Congr. Genet. Appl. Livest. Prod.*, Guelph, Ontario, Canada, 18:182.
 - 20 Strandén, I., and E. A. Mäntysaari. 1992. Animal model evaluation in Finland: experience with two algorithms. *J. Dairy Sci.* 75:2017.
 - 21 Swalve, H. H. 1995. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.* 78:929.
 - 22 Swalve, H. H. 1995. Test day models in the analysis of dairy production data—a review. *Arch. Anim. Breed.* 38:591.
 - 23 Van Tassell, C. P., R. L. Quaas, and R. W. Everett. 1992. Parameter estimates for 305-day ME records and 305-day test-day residual records. *J. Dairy Sci.* 75(Suppl. 1):251.(Abstr.)
 - 24 Wiggans, G. R. 1985. Procedures for calculating lactation records. *Natl. Coop. Dairy Herd Impr. Progr. Handbook. Fact Sheet G-1.* Ext. Serv., USDA, Washington, DC.
 - 25 Wiggans, G. R., and M. E. Goddard. 1996. A computationally feasible test day model with separate first and later lactation genetic effects. *Proc. N.Z. Soc. Anim. Prod.* 56:19.