# International Genetic Evaluation of Dairy Sires and Cows Using First Lactation Test Day Yields

J. Jamrozik<sup>1</sup>, L.R. Schaeffer<sup>1</sup> and K.A. Weigel<sup>2</sup>

<sup>1</sup> Centre for Genetic Improvement of Livestock, Department of Animal and Poultry Science University of Guelph, Guelph, ON, N1G 2W1, Canada <sup>2</sup> Department of Dairy Science, University of Wisconsin Madison, WI 53706, USA

# Introduction

International evaluation of dairy sires is currently a two-step procedure. First, within countries EBVs are estimated. Then, they are combined by the MACE method (Schaeffer, 1994) under a multiple trait sire model, where progeny performances in different countries are considered as genetically different traits. Genetic correlations between countries are estimated hv an approximate REML method with de-regressed sire EBVs as input data (Sigurdsson et al., 1996). International indices are influenced by several factors including within country methods of data recording and models for calculating EBVs. No direct estimates of across country cow breeding values are available.

A research project using individual animal performance records for international comparison was originated by K.A. Weigel in 2000. Data from seventeen countries was used to estimate genetic parameters from a multiple trait sire model, and lactation yields were analyzed in a multiple trait sire model for the purpose of international genetic evaluation (Weigel et al., 2001).

Test day (TD) models have been applied by several countries for genetic evaluation of dairy production traits. Accounting for the shape of an animal's lactation curve through regressions and modeling short-term environmental effects increases the accuracy of genetic evaluation (Swalve, 2000). Despite computational difficulties associated with multiple trait TD models, prospects exist for applying this methodology for international genetic evaluation of animals.

The objective of this study was to apply a multiple trait random regression TD model on data from selected countries.

# **Material and Methods**

#### Data

Data were first lactation TD milk yield (kg) records from four countries: Australia (AUS), Canada (CAN), Italy (ITA) and New Zealand (NZL). Countries were selected based on relatively small data sizes and a wide range of production systems and environments represented. Only milk yield in first lactation was analyzed. Cows of Holstein sires calving between January 1<sup>st</sup>, 1990 and December 31<sup>st</sup>, 1997 were included.

Selected countries differed in many aspects of their dairy production systems. Population sizes ranged from 0.9 mln (AUS) to 1.6 mln (NZL) first lactation cows, while average herd size ranged from 74 cows (CAN) to 97 cows (AUS). Average number of TD records per cow was smallest for NZL (3.9) and largest for ITA (8.9). Almost 100% ITA TD records came from supervised testing schemes, while only 11% and 19% were supervised in NZL and AUS, respectively. ITA and CAN had the highest level of daily milk production, followed by AUS and NZL. Sizable differences between persistency of lactation, measured as a difference in yields on DIM 280 and 60, existed between countries, with ITA having the most persistent overall lactation curve. Atypical shapes of AUS and NZL lactation curves late in the lactation (DIM>270) were attributed to smaller number of TD records in this interval.

Number of sires of cows with data ranged from 3007 (NZL) to 22416 (ITA). Only 173 sires had progeny in all four countries. This group of bulls included 113 sires with USA registration numbers, 57 Canadian bulls, and 3 sires with the Netherlands (NLD) as a country code in the registration number. Genetic links (through common sires) between countries were strongest between CAN and AUS with 657 common sires, and the weakest between ITA and NZL with 238 common sires.

Overall edits included: DIM between 5 and 305 days, age at calving between 18 and 38 months, daily milk yield between 0.1 and 100 kg. Only records corresponding to two times a day milking were used. In total, 32,303,346 TD records on 4,925,264 cows passed all edits.

Pedigree data for bulls were from the official Interbull pedigree file. Cow pedigrees were provided by the respective countries with data on sire, dam, maternal grandsire and maternal granddam included.

Parameters for the multiple trait random regression TD model were estimated using a data set created as follows:

- 1. Sires with daughters in all four countries (173 bulls) were selected,
- 2. Daughters of the 173 selected sires up to a maximum of 100 in a given country were kept,
- 3. Daughters of remaining sires, so that the total number of selected cows was between 14000 and 15000 per country, were added in random,
- 4. Records on dams (if any) of selected cows were added.

The resulting data set (556,168 TD records on 76,379 cows in total) is characterized in Table 1.

All data (after edits) was used to for genetic evaluation. Table 2 gives general characteristics of data, for both within and across country procedures. Genetic groups for unknown parents were based on sex, country of registration and birth year.

#### Model

The equation for single trait (within country) linear model was

$$y_{ijkt} = HTD_i + BSAxDIM_j + \sum a_{kl} z_{tl} + \sum p_{kl} z_{tl} + e_{ijkt}$$

where

 $y_{ijkt}$  is milk yield of cow k made on day t of lactation, within herd-test day effect i, for a cow belonging to subclass j for breed composition, age at calving, season of calving and DIM interval; HTD<sub>i</sub> is fixed herd-test day effect; BSAxDIM<sub>j</sub> is fixed breed composition, age at calving, season of calving and DIM interval effect;  $a_{kl}$  are random regression genetic coefficients specific to cow k;  $p_{kl}$  are random regression coefficients for permanent environmental (PE) effect on cow k;  $e_{ijkt}$  is residual effect for each observation;  $z_{tl}$  are covariates, assumed to be the same for genetic and PE regressions.

Regression curves were modeled by Legendre polynomials of order four. All covariates were scaled through multiplying by sqrt(2). DIM classes were formed as [5,20], [291,305], and 10 days intervals for DIM from 21 to 190. Classes were used in place of fixed regressions to model lactation curves. Residual variances were assumed constant within those intervals, all residual covariances were equal to zero.

The multiple trait (country) TD model combined the within country models through an appropriate covariance structure (additive genetic only) between countries.

# Methods

Bayesian methods with Gibbs sampling were used to estimate posterior means of covariance components for the multiple trait model. Single chain of 105,000 samples (with 10000 discarded) was generated. Gibbs sampling employed conditional normal distributions for systematic effects, multivariate normal distributions for genetic and PE effects, and inverted Wishart and Chi squared distributions for genetic and PE covariances, and residual variances, respectively.

Due to the orthogonal properties of Legendre polynomials, EBV for 305d yield (defined as sum of daily yields between days 5 and 305) is easily calculated as  $301*EBV(a_0)$ , while the genetic variance of 305d yield can be expressed as  $301^{2}*var(a_0)$ . Between country genetic correlations for 305d yields are the same as respective correlations for the genetic intercept  $(a_0)$ .

Genetic evaluations for both single and multiple trait models were done by standard mixed model methodology (Jamrozik and Schaeffer, 2000). Persistency of lactation was defined as 110\*[EBV(280)-EBV(60)].

# Results

#### Parameter estimation

Multiple trait estimates of variances for genetic random regression coefficients are presented in Table 3. Estimates were in relatively good agreement with corresponding single trait values (results not presented). An increase in the variance of the intercept over single trait estimates could be observed for AUS, CAN and NZL. Variances of other terms generally decreased in comparison with single trait estimates. PE regression variances (Table 4) were in general larger than genetic components. Heritabilities of intercept ranged from 0.38 (CAN and ITA) to 0.51 (AUS). Linear term (a<sub>1</sub>) had a lower heritability, from 0.21 (AUS) to 0.33 (CAN).

Posterior means of multiple trait estimates of genetic correlations between countries for the same lactation curve coefficients are shown in Table 5. Correlations ranged from 0.83 (intercept between AUS and NZL) to 0.17 (quadric term between CAN and NZL). The largest values of correlations were for the intercept with the smallest value of 0.66 between ITA and NZL. Correlations for the constant term were equivalent to correlations between 305d yields. The values were generally much smaller than one would anticipate based on earlier estimates from sire model using lactation yields (Weigel et al., 2001). The linear term (persistency component) had the lowest correlation between ITA and NZL. followed by CAN and NZL. CAN and ITA showed the highest genetic correlation for the persistency component.

#### Genetic evaluation

Estimated breeding values for 305d yield and lactation persistency (P) were calculated using EBVs of parameters of genetic lactation curves. Animal solutions of mixed model equations were used, without further scaling to any predefined genetic base. Correlation between 305d yield EBVs of sires estimated by single and multiple trait models for bulls with TD records on daughters ranged from 0.96 for NZL to 0.99 for CAN and ITA (Table 6). Correlations for persistency were slightly smaller (0.91 - 0.98). Sires with no daughters in foreign countries had greater correlations: 0.99 - 1.0 for 305d yield, 0.96 - 0.99 for P. Correlations for cows with TD records (Table 7) were similar in magnitude to sire correlations.

Across country correlation between sire (with daughters in single trait model) EBVs for 305d yield and persistency from single and multiple trait model are in Table 8. Except for the ITA – NZL pair, single trait correlations for 305d yield were smaller than estimated genetic correlations from multiple trait model. Correlations from multiple trait model were 0.96 (AUS-CAN, AUS-ITA, CAN-ITA and AUS-NZL) and 0.93 for CAN-NZL and ITA-NZL. Correlations between persistency EBV from single trait model were very low (0.11 - 0.53), higher for multiple trait model (from 0.17 for ITA-NZL to 0.85 for CAN-ITA). Across country correlations between cow (with TD records) EBVs from multiple trait model are presented in Table 9.

# Descussion

International evaluation of dairy sires using performance records could be the next step in the development of across country comparison methods (Weigel et al., 2001). Test day models are becoming more popular for within country genetic evaluation, so TD data will soon be equivalent to performance data for dairy production traits. Random regression models give tremendous opportunity for analysis of all aspects of lactation curves, including total yield and persistency, and differences in average length of lactation. Current computational restraints for random regression models should become less problematic in future due to technological progress in computer hardware.

The current study was the first attempt to apply a random regression TD model in the multicountry scenario. Several simplifying assumptions were made to facilitate computations. The same models were used for each of four participating countries. This could be a subject for improvement allowing for specific within country effects in the model. Estimation of across country covariances required a subset of the data to be used, that further reduced opportunities for a correct modeling (e.g. small number of observations for certain levels of fixed effects). Both genetic and PE regressions were modeled by the same general function (fourth order Legendre polynomials) for each country and a maximum lactation length of 305 days was assumed.

Parameter estimation with multiple country model was highly dependent on the quantity and quality of genetic ties between countries (Sigurdsson et al., 1996). An animal model was applied in this study which did not provide direct genetic links between countries (no cow had TD records in more than one country). All genetic ties were realized through sires with daughters in multiple countries. Selection of appropriate subsets of data for parameter estimation had to between balance data size and genetic connections.

Estimates of genetic correlations between four analyzed countries were much lower than reported elsewhere (Weigel et al., 2001). Reasons for this might be real differences in the genetics of lactation curve parameters between environments (=countries). Lack of sufficient genetic ties might also contribute to possible underestimation of genetic correlations. Further, our study used an animal model in contrast to sire models applied by others. Classical theory, however, advises that the same model should be used for both parameter estimation and genetic evaluation when covariances need to be estimated.

While the random regression TD model was applicable to first lactation records from four countries, there could be severe computational limitations if all 25 countries participating in the Interbull evaluation decided to do this. The total number of test day records could reach close to a billion or more, and the number of animals could be very large. Therefore, this approach to international dairy cattle comparisons may not be feasible with all countries for many years. Another drawback is that only first lactation milk yield records were used in this study, and practically all countries include records from all or several lactations for milk, fat, and protein yields in their national genetic evaluations. This study has not proven that the evaluations of bulls from these analyses are more or less accurate than existing MACE evaluations.

Given the short lactations in NZL and AUS compared to CAN and ITA, production yield in a lactation may be better expressed in terms of the average lactation length in a country rather than as 305-d yields. TD records between 5 and 350 days in milk should perhaps be included in genetic evaluation rather than to restrict these to between 5 and 305 days. The definition of persistency may need refinement for countries with short lactation length.

The international TD model has also provided EBVs for cows. The value of international rankings of cows has yet to be measured. The best cows are often discovered very quickly within a country and utilized to produce young bulls for AI. Thus, it is unlikely that these evaluations will uncover any cows that have not already been discovered. International access to these cows may also be limited. However, AI organizations that have offices in several countries may be able to make use of cows identified by these models. Dams of bulls are often required to have several lactation records (not just first lactation) before they are used to produce a young bull, and therefore, the TD model EBVs may be useful to spot potential bull dams if the EBVs are up-to-date for cows.

An international TD model is just another approach to ranking dairy bulls between countries and a full discussion of the advantages and disadvantages is necessary prior to planning the next steps in this or future projects.

# Acknowledgements

Data were provided by the genetic evaluation centres of Australia, Canada, Italy and New Zealand, and by Interbull Centre, Uppsala, Sweden. The authors are grateful to the Ontario Ministry of Agriculture, Food and Rural Affairs, DairyGen, and the Natural Science and Engineering Research Council of Canada for their financial support.

# References

Jamrozik, J. & Schaeffer, L.R. 2000. Comparison of two computing algorithms for solving mixed model equations for multiple trait random regression test day model. *Livest. Prod. Sci.*, 67, 143-153.

- Schaeffer, L.R. 1994. Multiple-country comparison of dairy sires. J. Dairy Sci. 77, 2671-2678.
- Sigurdsson, A., Banos, G. & Philipsson, J. 1996. Estimation of genetic (co)variance components for international evaluation of dairy bulls. *Acta Agric. Scand.* 46, 129-136.
- Swalve, H.H. 2000. Theoretical basis and computational methods for different test-day

genetic evaluation methods. J. Dairy Sci. 83, 1115-1124.

Weigel, K.A., Rekaya, R., Zwald, N.R. & Fikse, W.F. 2001. International genetic evaluation of dairy sires using a multiple-trait model with individual animal performance records. J. Dairy Sci. (Submitted).

Table 1. Data characteristics: multiple trait parameter estimation

Country	Records	Cows	Dams	Herds	HTD
AUS	141,206	14,315	3,781	3,372	79,243
CAN	147,745	14,041	4,315	5,232	88,979
ITA	177,434	14,170	7,740	2,601	84,638
NZL	89,783	14,696	4,085	2,695	14,844
Total	556,168	57,222	19,921	13,900	340,385

Table 2. Data characteristics: genetic evaluation

Country	Records	Cows	Animals	HTD
AUS	6,957,903	937,749	1,288,199	435,765
CAN	8,468,392	1,042,567	1,479,875	961,077
ITA	11,087,435	1,385,257	1,973,459	1,030,889
NZL	5,789,616	1,559,691	2,101,401	252,410
Total	32,303,346	4,925,264	6,838,077	2,680,141

Table 3. Posterior means (standard deviations in parenthesis) of multiple trait estimates of variances of genetic random regression coefficients

Country	$\mathbf{a}_0$	<b>a</b> <sub>1</sub>	$\mathbf{a}_2$	a <sub>3</sub>	$\mathbf{a}_4$
AUS	3.25(0.26)	0.22(0.04)	0.10(0.01)	0.07(0.010)	0.03(0.005)
CAN	5.10(0.55)	0.68(0.09)	0.28(0.04)	0.14(0.019)	0.06(0.010)
ITA	6.02(0.47)	0.51(0.06)	0.34(0.04)	0.12(0.018)	0.06(0.010)
NZL	1.99(0.14)	0.24(0.03)	0.09(0.01)	0.03(0.005)	0.02(0.004)

Table 4. Posterior means (standard deviations in parenthesis) of multiple trait estimates of variances of PE random regression coefficients

Country	$\mathbf{a}_0$	<b>a</b> <sub>1</sub>	$\mathbf{a}_2$	<b>a</b> <sub>3</sub>	$\mathbf{a}_4$
AUS	3.13(0.19)	0.80(0.04)	0.33(0.03)	0.19(0.015)	0.14(0.011)
CAN	8.22(0.46)	1.36(0.10)	0.64(0.05)	0.29(0.034)	0.23(0.023)
ITA	9.95(0.40)	1.85(0.08)	0.71(0.05)	0.31(0.028)	0.25(0.022)
NZL	2.20(0.10)	0.63(0.03)	0.26(0.02)	0.15(0.010)	0.11(0.008)

Table 5. Posterior means (standard deviations in parenthesis) of multiple trait estimates of between countries genetic correlations for random regression coeffcients

	AUS-CAN	AUS-ITA	AUS-NZL	CAN-ITA	CAN-NZL	ITA-NZL
$\mathbf{a}_0$	0.79(0.04)	0.74(0.04)	0.83(0.03)	0.77(0.04)	0.72(0.04)	0.66(0.05)
<b>a</b> <sub>1</sub>	0.43(0.09)	0.42(0.09)	0.39(0.08)	0.55(0.08)	0.28(0.10)	0.16(0.08)
<b>a</b> <sub>2</sub>	0.61(0.08)	0.59(0.07)	0.23(0.11)	0.59(0.09)	0.21(0.11)	0.03(0.11)
<b>a</b> <sub>3</sub>	0.52(0.06)	0.47(0.09)	0.43(0.11)	0.46(0.09)	0.44(0.11)	0.50(0.10)
$\mathbf{a}_4$	0.35(0.12)	0.46(0.09)	0.21(0.12)	0.37(0.12)	0.17(0.12)	0.27(0.12)

and instantion persistency (1) (not in animoti of daughters, 1(12) no daughters on a foreign search)							
	Bulls	AUS	CAN	ITA	NZL		
305 d yield	Nod>0	0.98	0.99	0.99	0.96		
	nod>19, NFD	1.00	0.99	0.99	0.99		
Р	Nod>0	0.95	0.98	0.98	0.91		
	nod>19, NFD	0.97	0.99	0.99	0.96		

Table 6. Correlations between EBVs of sires estimated by single and multiple trait models, for 305 milk yield and lactation persistency (P) (nod = number of daughters, NFD = no daughters on a foreign scale)

Table 7. Correlations between EBVs from single and multiple trait models, for 305 milk yield and lactation persistency (P) on cows with TD records

	AUS	CAN	ITA	NZL
305 d yield	0.99	0.99	1.00	0.99
Р	0.94	0.98	0.96	0.93

Table 8. Across country correlations between sire EBVs (nod>0 in single trait model) for milk yield and lactation persistency (P), from single and multiple trait models

•	• • •	AUS-CAN	AUS-ITA	AUS-NZL	CAN-ITA	CAN-NZL	ITA-NZL
305d yield	Single	0.67	0.69	0.74	0.75	0.64	0.71
-	Multiple	0.96	0.96	0.96	0.96	0.93	0.93
Р	Single	0.19	0.13	0.53	0.50	0.11	0.19
	Multiple	0.51	0.33	0.73	0.85	0.37	0.17

Table 9. Across country correlations between cow EBVs (with data) for milk yield and lactation persistency (P), from multiple trait model (n=4925264)

	AUS-CAN	AUS-ITA	AUS-NZL	CAN-ITA	CAN-NZL	ITA-NZL
305d yield	0.97	0.96	0.97	0.97	0.94	0.93
Р	0.48	0.37	0.70	0.86	0.26	0.15