# The Effect of Selection and Data Structure in National and International Evaluation: A Simulation Study

***Canavesi, F.***
*ANAFI, Italy*

## Introduction

De-regressed proofs (DRPF) are the recommended variable in international genetic evaluation procedure (Sigurdsson, 1995). Different analysis on Italian data did show a decreasing variability trend in DPRF standard deviation over time (Cassandro, 1996) that had an impact on sire variance estimates. Successively time editing was applied to data to avoid that trend (increasing or decreasing) over time of DRPF standard deviations (SD) greatly influenced the ability of Mace to correctly compare more recent sires with the old proven bulls within a particular country (Weigel, 1996).

The decreasing trend found in Italy, greater in change compared to other countries were thought to be due to the selection programmme that was changing over time along with number of bulls proven each year and average number of daughters per bull.

Very recently a significant difference in trend was found in the Italian population between SD of daughter-yield-deviations (DYD) and SD of DRPF over time, the latest dependent variable constantly producing a decreasing trend over time (Canavesi, 2000).

The objective of this study was to test in a simulation the effect of data structure and selection:

1. on SD over time of DRPF compared to SD of DYD;
2. on sire variance estimates
3. on accuracy of estimated breeding values.

## Material and Methods

The complete Italian lactation data set of February 1999 was re-simulated generating for each record in the file a new lactation.

Production records were simulated as:

$$Y = \mu_{HY} + TBV_{anim} + PE_{anim} + error$$

where

| | |
|---|---|
| $\mu_{HY}$ | is the real within herd average milk production varying according to year of production. |
| $TBV_{anim}$ | is the true breeding value of the animal |
| $PE_{anim}$ | is the permanent environmental effect of each cow. |

A phenotypic standard deviation of 1,200 kg was simulated with an heritability of 0.25 which sets the genetic standard deviation to 600 kg.

For each ancestor in the base population (animals with unknown parents) the TBV was randomly generated from a normal distribution with mean 0 and a SD of 600 kg. The progeny TBV were then simulated as:

$$TBV_{anim} = ½TBV_{sire} + ½TBV_{dam} + MS$$

where

$TBV_{sire}$ is the breeding value of the sire
$TBV_{dam}$ is the breeding value of the dam
MS is the mendelian sampling

Data were analyzed separately with the official programs used for the Animal Lactation Model with repeated records currently used in Italy. The evaluation for both data sets included a maximum of three lactations per cow. Heritability was in both cases 0.25.

Ten replicates were simulated.

Animal Model EBV were correlated with simulated TBV.

DRPF proofs were then computed for each replicate using Interbull programs and compared with DYD from the Animal Model evaluation.

Sire standard deviations estimates were also compared to the expected value of 300 kg .

Data structure in terms of number of daughters and contemporary size over time may play a role in creating trends, since number of daughters and lactation in the Italian bull popuplation is not constant over time. The new weighting factors presently undergoing test were applied to data to verify the behaviour of DRPF over time using weights that better account for data structure.

## Results and Discussion

A total of 4,875,538 lactation records were generated related to 2,806,684 cows. A total of 3,510,824 animal effects were included in the simulated data. Calving dates ranged from January 1983 to Novermber 1998. Average herd-size was 40-50 animals per herd-year.

A total of 5408 bulls were evaluated with at least 10 daughters in 10 herds (Figure 1).

Since no selection was simulated bull genetic trend was 0 randomly fluctuating over time due to data structure and random high use of a particular genetic value (Figure 2).

Standard deviation of DYD did not show any trend, while DRPF had a significant decreasing trend over time with a slope on average (over 10 replicates) of 2.227 kg per year (Figure 3). Thus DYD and DRPF in an Animal Model with repeated record and a complex data structure were not exactly equivalent. Estimated sire variances, equal on average to 307.9 kg were constantly slightly higher than expected; a time editing on birth year of bulls did reduce the overestimation to 303.7 kg showing a significant effect of the trend on estimated sire variances.

Correlation (simple and rank) between Estimated and true breeding value was fairly costant around .93.

Use of alternative weighting resulted in an increased trend over time and in a reduced overstimation of the sire variance that resulted equal to 305.5 kg (Figure 4). The slope of the decreasing trend was equal to –10.863.

## Conclusions

DYD and DRPF do not behave equally over time in a real population with repeated record, DYD were more cosntant over time than DDRP. In a contest were selection was not simulated it is perhaps necessary to check whether the decreasing trend may have an effect on international ranking of bulls. The effect on sire estimated standard deviation was small but it was present.

Second step of the simulation will be to verify what happens when selection is on place.

## References


1. Canavesi, F. & Samorè, A.B. 2000. Phantom groups and equivalence between daughter yield deviations and deregressed proofs as critical elements in MACE International evaluations. *Interbull Meeting, Bled,* 14-15 May 2000.
2. Cassandro, M., Canavesi, F., Brandts, A., Carnier, P., Gallo, L., Bittante, G. & Bagnato, A. 1996. Trend of within country sire variance and potential impact on international evaluations for production traits. *Interbull bulletin no. 14.*
3. Schaeffer, L.R. 1994. Multiple-country comparison of dairy sires. *J. Dairy Sci. 77,* 2671-2678.
4. Sigurdsson, A. & Banos, G., 1995. Dependent variable in International Sire Evaluation. *Acta Agric. Scand. 45,* 209-217.
5. Van Raden, P.M. & Wiggans, G.R. 1991. Derivation, calculation and use of national animal model information. *J. Dairy Sci. 74,* 2737-2746.
6. Weigel, K. & Banos, G. 1996. Effect of time period of data used in international dairy sire evaluation. *Interbull workshop, Verden.*

Figure 1. Distribution of bull proofs.

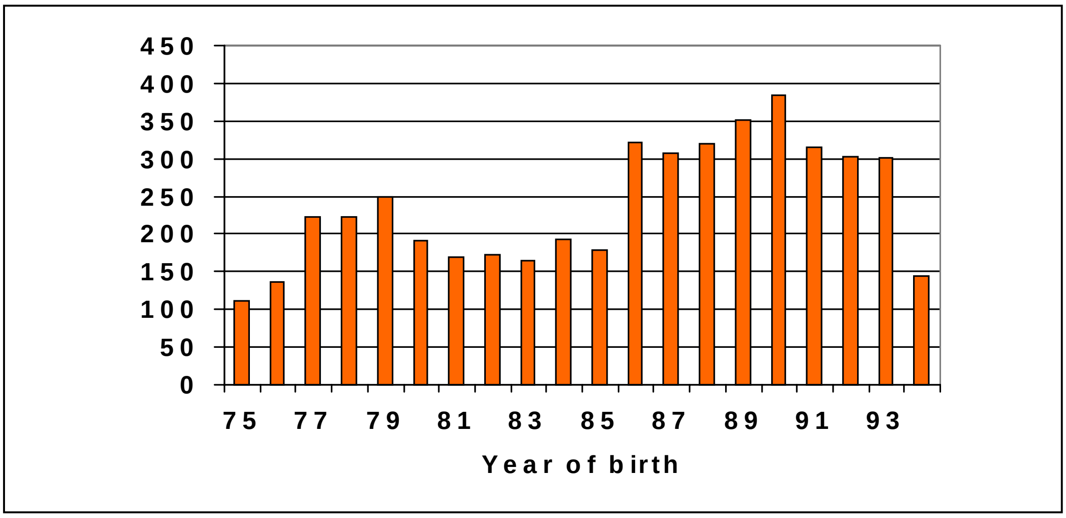


Figure 2. Bull genetic trend.

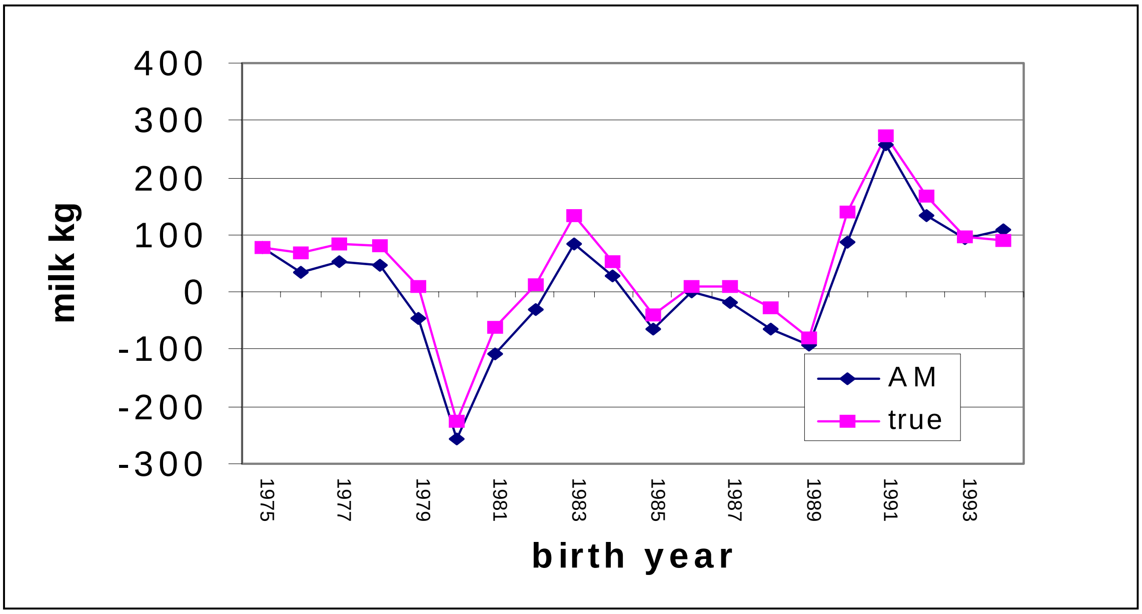

Figure 3. DYD and DRPF, using number of daughters as weight for de-regression.

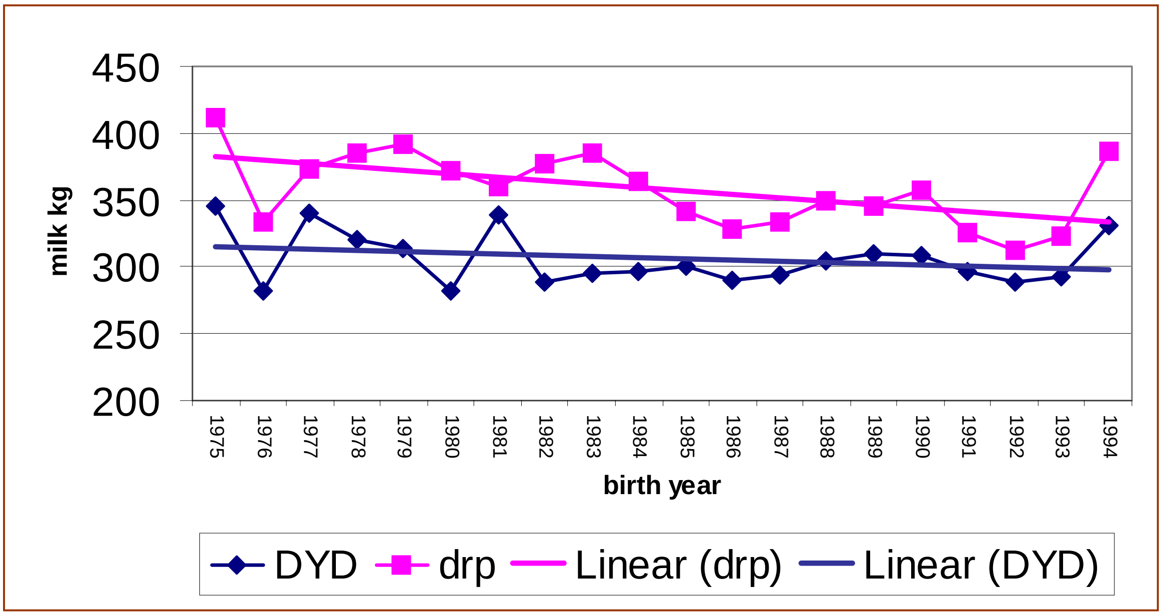


Figure 4. DYD and DRPF, using number of daughters as weight for de-regression.