

A Comparison Between Single and Composite Samples for Dairy Cattle Evaluation

*F.W.C. Naser^{1#}, R van Dyk¹ F.H. Kanfer² & *J. v.d. Westhuizen³*

¹ Department of Animal Science, University of the Free State, PO Box 339, Bloemfontein, 9300, Republic of South Africa. ² Department of Statistics, University of Pretoria, Pretoria, 0002, Republic of South Africa. ³ ARC-Animal Improvement Institute, P Bag X2, Irene, 0062, Republic of South Africa.

Abstract

A simulation study was carried out to compare the use of single and composite samples in dairy cattle evaluation. Multiple trait analyses were carried out to evaluate the difference between a single and composite sample. Results showed a high genetic correlation between the breeding values of the composite and single sample. This indicates that the same genes are responsible for both traits and that it is possible to make use of a single sample instead of a composite sample.

Keywords: Butterfat, composite sample, Dairy cattle evaluation, single sample

Introduction

Milk recording entails measuring daily milk yield and composition and is used for herd management decisions as well as cow and sire evaluations. It was first implemented in South Africa in 1917, by the then Friesian Breed Society and subsequently taken over by the State in 1919 (Grobler & Loubscher, 1983). The Milk Recording Scheme was run entirely by the State and included on-farm tests to check that participants followed the correct procedure. During 1975 it was decided to simplify the Scheme and allow farmers to participate more actively in the collection of multiple samples (morning and evening), which were sent to a central laboratory. In the laboratory the two samples were mixed and then analysed. Some errors did occur in the collection of the samples, but was compensated for by what was thought to be a more accurate method of analysing in a central laboratory (Annual Report - ARC, 1996).

In order to make milk recording user-friendlier the possibility of replacing composite sampling with a single sample (1995/6) was investigated. Various reasons were presented for this simplification, the most important obviously being financial benefits and user-friendliness. The implementation of the new Scheme was based on the fact that there were practically no differences

between two test years averages (one year composite - and second year single sample). This comparison is, however, not scientifically validated, since there were no data available to make a scientific evaluation and compare the differences. In order to compare the two sampling methods, it is important to take both samples (single and composite) at the same time and day. A test day model for genetic evaluation can account for factors that are specific to each test day, such as management groups within a herd on test day, day of the year, and for each cow, days in milk, pregnancy status and number of times milked on test day. This clearly points to the problem of data collection. Many of these factors change from one test day to the next and would be difficult to model for 305 days yield (Jamrozik *et al.*, 1997). Errors are one third larger when estimates of lactation milk yield were based on one milking, one day each month rather than on both milkings (Smith & Pearson, 1981).

The gain in accuracy in prediction for one trait from using other correlated traits is partitioned into a direct gain from measuring other traits and gain because fixed effects are estimated more precisely (Thompson & Meyer, 1986). Even slight improvement of accuracy can have dramatic economic effects in large populations (Pollak *et al.*, 1984). The aim of this study was to determine whether the decision to replace the composite sample with a single sample is justified. This was

done by comparing simulated composite and single samples in a multiple trait analysis.

Materials and Methods

To obtain information on the advantage and accuracy of having a single, instead of composite sample, records are needed. Suitable records are scarce since the collection is extremely time consuming and costly. De Waal (2001) did an investigation on the influence of sample frequency during milk recording on the reliability of performance testing in dairy cows. Single and composite samples were taken in four herds. A single sample was taken in the morning and evening at the same time as the two samples comprising the composite sample. Records were obtained from 236 cows.

Because of the lack of “real” data the information from this study was used to simulate data in order to investigate this problem. The information is used to quantify the relationship between a single and composite sample.

Only butterfat was used in this study in order to simplify the calculation. Butterfat is historically defined as the amount of butterfat in a 36ml milk sample taken during milking over a period of 24 hours. A single sample will change the definition to the amount of butterfat found in a sample taken after 12 noon on the test day. From the resulting measurements of butterfat regarding the two methods, the single sample will be calculated according to the differences between the two measurements. The question that now arises is: Can accurate breeding values for butterfat still be predicted?

It was furthermore decided to only simulate a single and composite sample under a selection scenario. The effect of selection was discussed by Naser *et al.* (2001). Twenty-five repetitions for each scenario was simulated in order to keep the analysis as simple as possible.

Simulation

In the first four years all females generated were retained, where after the herd operated as a normal dairy enterprise. One hundred AI sires, all from the same genetic level, were simulated. In the first year ten bulls were used as sires, in the second

year five new bulls were added while five bulls from the previous year were retained. This pattern continued for 20 years and was done to establish strong genetic ties between years. This means that bulls were used on a random, non-selected, basis. This process was repeated on an annual basis for the remaining term. For each simulated cow in the herd the following information was recorded: cow number, fixed effect level at first lactation, production measurement at first lactation, genetic component of the measurement, error component of measurement, sire and dam. These animals were replaced as described by the selection scenario.

Selection scenario: Animals were ranked according to the genetic component of their butterfat measurement. The 25% poorest performing animals were replaced by the best performing progeny. Cows older than six years were also replaced by the remaining best performing progeny. Selection was done on a yearly basis.

Measurements were simulated applying a model consisting of a fixed effect, random genetic and random error component. Only these components were included to keep the model simple. For the purpose of this study the fixed effect can be viewed as a year effect, or perhaps a herd-year effect.

The model used to simulate the composite sample can be written as:

$$y_{ij} = f_i + a_{ij} + e_{ij}$$

where

- f_i represents the fixed effect at level i ,
- a_{ij} represents the random genetic component of animal j under fixed effect i ,
- e_{ij} represents the random error component,
- y_{ij} represents butterfat.

The random genetic component- (a_{ij}) and error component (e_{ij}), have normal distributions with average 0 and variances σ_a^2 and σ_e^2 respectively. (($a_{ij} \sim N(0, \sigma_a^2)$), ($e_{ij} \sim N(0, \sigma_e^2)$)).

This corresponds to the methods used by Van Vleck (1993). He described the method of simulation as to obtain, in some way, pseudo-random values from a normal distribution with a mean, zero, and variance, one. This is similar to a

Monte Carlo simulation. Tuchscherer & Herrendörfer (1998) and Canavesi & Miglior (1999) proposed similar models to evaluate estimated BLUP:

The fixed effect component is calculated by a prescribed formula defined as:

$$f_i = c + m(i)$$

with:

$$\begin{aligned} m_i &= 12.5 \text{ and } c_i = 100 \text{ if } 1 \leq i \leq 8, \\ m_i &= -4.5 \text{ and } c_i = 201 \text{ if } 9 \leq i \leq 16 \text{ and} \\ m_i &= 10 \text{ and } c_i = 10 \text{ if } 17 \leq i \leq 20. \end{aligned}$$

The genetic and error component were obtained by generating random numbers from a normal distribution with mean 0 and variances σ_a^2 and σ_e^2 respectively. The genetic- ($\sigma_a^2 = 293$) and error variance ($\sigma_e^2 = 534$) from the study of du Toit *et al.* (1998) was used in the simulation. This implies that a heritability estimate of 0.35 was used.

The following model simulated the single sample:

$$y_{ij} = f_i + a_{ij} + d_{ij} + e_{ij}$$

with

- f_i represents the fixed effect at level i ,
- a_{ij} represents the random genetic component of animal j under fixed effect i ,
- e_{ij} represents the random error component,
- y_{ij} represents butterfat, and with the genetic and error components statistically independent.
- d_{ij} a random difference between composite and single sample as calculated from the data of de Waal & Heydenrych (2001).

Multiple trait analyses were done using REML (Meyer, 1998) procedures, for all 25-simulation rounds to determine the (co)variance components and subsequent genetic correlation estimate between the single and composite samples.

Starting values from the single trait analysis were used for the analysis of the first simulation round. Thereafter results of the first simulation were used as starting values for the remaining simulation rounds. Because of the high genetic correlation that exists between the two traits it was

difficult to reach convergence using the simplex search method, the Powell (1965) search method was, therefore, used.

Results and Discussion

In the first four years an underestimation of the fixed effect in both traits occurred. No selection took place during this period and all animals were retained. After this period (year 5-20) an overestimation of the fixed effect occurred. This could be ascribed to the selection that took place during that period. It is interesting to note that there was virtually no difference between the estimated fixed effect levels of the two traits.

The mean variance component estimates (for 25 rounds of simulation) for the single and composite samples are as follows:

Single sample heritability estimate	16.70
Single sample additive variance	81.37
Single sample error variance	407.34
Composite sample heritability estimate	19.16
Composite sample additive variance	100.02
Composite sample error variance	422.30

Both the additive and error variance in the composite samples was higher than those obtained in the single samples. This also reflected in the heritability estimates. It should be noted that the true breeding value of an animal was used as basis for the simulation of the single and composite sample and this will explain the similarities of additive variance estimates between different simulation rounds. In each case the genetic correlation between the two traits were estimated as unity.

Conclusion

The results from the multiple trait analysis showed that although individual variance estimations differed, a perfect positive (unity) genetic correlation exists between the single and composite sample. This study showed that it is possible, in practice, to make use of a single sample instead of a composite sample in order to determine breeding values for dairy cattle. This is important in order to have a simplified scheme, which is easy and more economic to manage.

References

- Annual Report. 1996. National Dairy Cattle Performance Testing Scheme: South Africa. *Agriculture research council, Irene, South Africa*.
- Canavesi, F. & Miglior, F. 1999. Effect of De-Regression on Lactation Number on MACE Evaluations: A Simulation study. *In: Proc. of the Interbull Meeting Zurich, Switzerland*. 22, 49-51.
- De Waal, H. & Heydenrych, H.J. 2001. The effect of sampling frequency on the accuracy of estimates of milk-fat yields of dairy cows. *S. Afr. J. Anim. Sci.* 31, 9-11.
- Du Toit, J., van Wyk, J.B. & van der Westhuizen, J. 1998. Genetic parameter estimates in the South African Jersey breed. *J. Anim. Sci.* 28, 146-152.
- Grobler, B.R. & Loubser, L.F.B. 1983. Die ontwikkeling van melkprestasie toetsing in die RSA. *Nasionale Melkprestasie-en-Nageslag Toetsskema. Jaarverslag*, 3-6.
- Jamrozik, J., Schaeffer, L.R. & Dekkers, J.C.M. 1997. Genetic evaluation of dairy cattle using test day yields and Random Regression Model. *J. Dairy Sci.* 80, 1217-1226.
- Meyer, K. 1998. DFREML version 3.0. User notes.
- Neser, F.W.C., Van Dyk, R. & Kanfer, F.H. 2001. The effect of selection on genetic parameter estimates. *S. Afr. J. Anim. Sci.* (Submitted).
- Powell, M.J.D. 1965. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comp. J.* 7, 155-162.
- Pollak, E.J., van der Werf, J. & Quaas, R.L. 1984. Selection bias and multiple trait evaluation. *J. Dairy Sci.* 67, 1590-1595.
- Smith, J.W. & Pearson, R.E. 1981. Development and evaluation of alternate testing procedures for Official Records. *J. Dairy Sci* 64, 466-474.
- Thompson, R. & Meyer, K. 1986. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest. Prod. Sci.* 15, 299-313.
- Tuchscheherer, A. & Herrendörfer, G. 1998. Evaluation of estimated BLUP in mixed linear models by a designed computer simulation. *Proc 6th World Congr. Genet. Appl. Livest.* 25, 585-588 (Armidale Australia).
- Van Vleck, L.D. 1993. *Selection index and introduction to mixed model methods*. CRC Press Inc. Florida.