Estimation of fat and protein yields in dairy cattle from one milk sample per day in herds milked twice a day

M. Trinderup, T. Lykke & U. Sander Nielsen

Danish Agricultural Advisory Service, National Centre, Danish Cattle Federation, Udkaersvej 15, Skejby, 8200 Aarhus N., Denmark

Summary

A linear regression model was developed to estimate the daily yield of kg fat and protein from milk samples taken either at morning or evening milking in herds milked twice a day. The recommended model includes information on breed, parity, classes of days from calving, milk yield (kg) morning and evening, fat and protein yield (kg) at one milking and time interval between the milkings. The time interval was obtained from the electronic milk meters. A modified model was developed including the milk yield relation morning and evening as a measure of the distance between milkings. This model can be used in herds where the time of the milking is not recorded automatically. The two models were estimated on one half of a dataset that included information on 63 077 cows and tested on the other half of the data. Accuracies for kg protein obtained on the test data were for morning milk samples for the three parities between 0.9910 and 0.9944 for the recommended model and between 0.9909 and 0.9943 for the modified model. For evening samples the accuracies were slightly lower. The accuracies for kg fat were between 0.931 and 0.944 for the three parities and the two models for morning milk samples and slightly larger for evening milk samples. The performance of the two models indicated that a testing scheme with one milk sample per test day can be applied in herds without electronic milk meters if the time interval is substituted with the milk relation. For 4 418 Danish Holstein cows the 305 days yield were calculated using four different sampling methods. If the milk samples were strictly morning or evening samples the accuracies were not as high as if the samples were taken alternately morning and evening throughout the lactation.

***Key words**: Testing schemes, Daily yield, One milk sample.*

Background

Danish herds increase and the average herd size at present is 125 cows. The farmers ask for new systems that can reduce the amount of hours and costs used in relation to the milk recording system. To improve the milk recording system the Tru-Test

electronic milk meters were introduced to herds in Denmark. But doing this the costs were increased from one milk sample per cow per test day to one milk sample pr milking per cow per test day.

In Germany and Canada, however, a reduced number of milk samples per milk recording are used to calculate the daily yield of fat and protein. It has been a request from the Danish milk recording organisation to have more knowledge on the possibilities of reducing the number of milk samples.

For a test period from 2003 to 2004 the milk recording was carried out using Tru-Test electronic milk meters in a number of herds. Milk samples were collected and analysed per milking, providing a data set for developing a model to estimate the daily fat and protein yield from one milk sample.

With fewer milk samples we expect to have a lower accuracy of the calculated yield on the test day, but the question is to what extent will this affect the calculated lactation yield?

Some farmers milk the cows three times or more per day, both with traditional milking systems and with automatic milking systems. These procedures cause a new situation, which have to be handled. Up to now a small investigation has been made in relation to the traditional milking three times per day, but more work has to be done to solve the problems in relation to more milkings with irregular distances between milkings.

The purpose of this investigation was to develop a model and estimate the necessary parameters to calculate the daily yield of fat and protein on the basis of one milk sample per day. Given the results from the model it was investigated what the impact of a reduced number of milk samples per test day has on the lactation yield.

Material

The data were collected from June 2003 to July 2004 from 540 herds with two milkings per day. The data included 63 077 cows recorded and in total 478 180 milkings. The model was developed for the three main dairy breeds in Denmark but only results from 38 682 Danish Holstein cows are presented in this paper.

Table 1. Average results for cows of the Danish Holstein dairy breed.

Number: 151.773 "cow test days"					
	38 682 Cows	Mean	Std.	Min.	Max.
Milk, kg	Yield per day	27.01	7.98	2.28	72.30
	Morning	14.84	4.48	1.07	46.49
	Evening	12.16	3.90	1.00	43.81
	Difference a.m. – p.m.		2.68		
Fat, %	Yield per day	4.18	0.69	2.00	7.80
	Morning	3.95	0.75	2.00	8.00
	Evening	4.45	0.80	2.00	8.00
	Difference a.m. – p.m.	0.50			
Protein, %	Yield per day	3.42	0.37	2.12	7.30
	Morning	3.40	0.37	1.66	7.60
	Evening	3.44	0.38	1.59	9.09
	Difference a.m. – p.m.	0.04			
Time interval	p.m. – a.m. milking	13.38	0.78	8.29	17.30
	a.m. – p.m. milking	10.62	0.78	6.71	15.71

Morning milkings were accepted from 3 to 10 a.m. and evening milkings from 2 to 8 p.m. Fat percentages between 2 and 8 and protein percentages larger than 1.5 were accepted. The milking interval from p.m. to a.m. was given from the milking times recorded by the TRU-Test milk meters, while the interval from a.m. to p.m. was found by extracting the p.m.-a.m. interval from 24 hours.

At each milking the milk yield was recorded and a milk sample was obtained. The milk samples were analysed and the fat and protein content was given in percentages. The daily fat and protein percentages were found as the sum of the morning and evening yield in kg divided by the daily milk yield.

Data was divided into two datasets according to herds chosen at random. The model for the daily yield was estimated on the first dataset. The robustness of the model was then tested on the second dataset (the test data).

A third dataset was constructed containing data for the period June 2003 to January 2005 from herds with two milkings per day and 11 test days per year. This dataset included information on the lactations of 4 418 Danish Holstein cows. On these data it was tested what impact a reduced number of milk samples per test day has on the lactation yield.

Method

Schaeffer *et al*. (2000) recommends a model including information on milk yield from both milkings, the fat and protein yield from one milking, milking interval, days in lactation, calving number and calving season to estimate the daily yield of fat and protein. The model has a regression for each of 72 classes.

The Danish model was developed in the same way as the model from Schaeffer. Nine different models were analysed but only the recommended model is presented in this paper.

It is assumed that the fat yield per day (y_{F24}) depends linearly of the fat yield from one milking (y_{F1}). The models that were used to estimate the protein yield per day look like the models used for fat. The only difference was that y_{F24} was substituted by y_{P24}.

$$y_{F24} = \beta_0 + \beta_{1(ijk)} + \beta_{2(ijk)}y_{F1} + \beta_{3(ijk)}y_{M1} + \beta_{4(ijk)}y_{M2} + \beta_{5(ijk)}y_{P1} + \beta_{6(ijk)}t$$

where i=1 to 3 (breed), j=1, 2, 3(+) (parity), k=1 to 4 (classes of days from calving) and where t is the time interval between the milkings, y_{M1} is the milk yield at the milking where the milk sample is sampled, y_{M2} is the milk yield at the other milking on the test day, y_{P1} is the protein yield (kg) and y_{F1} is the fat yield (kg) measured at milking M1.

The model was estimated for the fat and protein yield expressed both in kg and in percentage.

From the literature it is known that the distance between milkings is important information in the yield calculation. Using the electronic milk meters the starting time for each milking is known, but using other types of equipment it can be missing information. Therefore it was analysed with a modified model what the value was of using the relation between kg milk (morning) and kg milk (evening) to describe the distance between milkings. In the modified model the time interval was substituted with the a.m./p.m. milk yield relation ($m_{AM/PM}$).

As a measure of the robustness of the model the accuracy was calculated on the test data. The accuracy is defined by Schaeffer *et al*. (2000) as

$$Acc = Var_{(act)} / (Var_{(act)} + Var_{(dif)})$$

where $Var_{(act)}$ is the variance of the actual 24 hour yield and $Var_{(dif)}$ is the variance of the difference between the actual and the estimated 24 hour yield. The more accurate the yield is estimated the smaller the variance of the difference and the closer to 1 Acc will be.

Schaeffer *et al*. (2000) found that for a milk sample taken in the morning the accuracy was 0.94 for kg fat and 0.995 for kg protein. A sample taken in the evening had a slightly lower accuracy.

Results

Table 2 shows R^2-value, the standard deviation of the estimate of the 24 hours' yield ($S.D._{y24}$) and the standard deviation of the residuals [$(MSE)^{½}$] for the model by estimation of the fat and protein yield per day in kg and percentage.

If R^2-values for kg fat and protein (Table 2) are compared with R^2-values for fat and protein expressed as a percentage, lower R^2-values are achieved for estimation of % fat and protein.

The R^2-values for estimation of kg fat is highest with the p.m. milk sample. For estimation of kg protein the R^2 is highest with a.m. milk samples. The following results are from the model where fat and protein yield in kg are estimated.

Table 3 shows the results from testing the recommended and the modified model on the test data. The table shows the mean difference between the known 24 hours' yield and the estimated yield defined as dif = known - estimated. If the difference is positive, the estimated values on average were lower than the known values and vice versa.

By comparing the accuracy of the recommended model with the modified model, it becomes clear that the modified model is almost as precise as the recommended model. This will as mentioned earlier be an advantage as one milk sample per cow per test day also can be applied in herds where Tru-Test milk meters are not used today without further disadvantages in the shape of manual registration of the starting time of the milking. The standard deviation of the differences ($S.D._{Dif}$) increases a little, though.

Table 2. R^2-value, the standard deviation of the estimate of the 24 hours yield ($s.d._{y24}$) and the residuals [$(MSE)^{½}$].

		Fat			Protein		
Milk Sample	R^2	$s.d._{y24}$	$(MSE)^{½}$	R^2	$s.d._{y24}$	$(MSE)^{½}$	
a.m.	Kg	0.9395	0.3011	0.0765	0.9940	0.2325	0.0181
	%	0.9277	1.0006	0.2797	0.9823	0.4659	0.0625
p.m.	Kg	0.9432	0.3017	0.0741	0.9935	0.2319	0.0188
	%	0.9123	0.9923	0.3081	0.9751	0.4642	0.0743

Table 3. The accuracy for the recommended and the modified model on the test data.

		Recommended model (time interval)			Modified model (a.m./p.m. milk yield relation)		
	Parity	Mean dif.	$s.d._{Dif}$	Acc	Mean dif.	$s.d._{Dif}$	Acc
Kg fat							
a.m.	1	0.0035	0.0662	0.9329	0.0043	0.0671	0.9312
	2	0.0042	0.0830	0.9430	0.0058	0.0848	0.9406
	3	0.0023	0.0911	0.9438	0.0037	0.0928	0.9417
p.m.	1	-0.0028	0.0632	0.9385	-0.0037	0.0637	0.9376
	2	-0.0040	0.0786	0.9485	-0.0054	0.0794	0.9476
	3	-0.0035	0.0860	0.9496	-0.0048	0.0869	0.9485
Kg protein							
a.m.	1	-0.0007	0.0179	0.9910	-0.0005	0.0180	0.9909
	2	-0.0006	0.0202	0.9938	-0.0003	0.0204	0.9937
	3	-0.0008	0.0206	0.9944	-0.0005	0.0208	0.9943
p.m.	1	-0.0010	0.0192	0.9897	-0.0009	0.0193	0.9895
	2	-0.0009	0.0207	0.9935	-0.0008	0.0207	0.9935
	3	-0.0012	0.0207	0.9944	-0.0011	0.0208	0.9943

For fat the accuracy is highest with the p.m. samples. For protein the accuracy is highest with the a.m. samples. Therefore, indications are that there in average won't be any difference whether the milk samples are taken out in the evening or in the morning and therefore practical considerations may determine when the samples are taken out.

Estimation of 305 days' yield

Using the recommended and the modified model and only one milk sample per test day the daily yield was estimated on the third dataset. From the estimated daily yields the 305 days yields were calculated. To select which milk sample per test day that had to be used to estimate the daily yield four different sampling methods was developed. Table 4 shows the four methods for selecting the milk samples per test day that the 305 days' yields are based on.

Table 5 shows the accuracy of 305 days' yields in kg fat and kg protein estimated according to the four methods shown in table 4.

Table 4. View of the combinations of time of day when the milk sample is obtained.

Name	Time of day for the milk sample
AM	All samples through the lactation were morning samples
PM	All samples through the lactation were evening samples
AM1PM2	Every second sample was a morning sample and every other sample was an evening sample. The first test day after calving the sample was a morning sample.
AM2PM1	Every second sample was an evening sample and every other sample was a morning sample. The first test day after calving the sample was an evening sample.

Table 5. Accuracy of 305 days yield depending of method of milk sampling.

		Parity	Number	AM	PM	AM1PM2	AM2PM1
Recommended model (time interval)	Fat	1	2 368	0.96	0.94	0.98	0.97
		2	967	0.95	0.93	0.97	0.96
		3	1 083	0.95	0.93	0.97	0.96
	Protein	1	2 368	0.998	0.998	0.998	0.999
		2	967	0.998	0.998	0.999	0.999
		3	1 083	0.999	0.999	0.999	0.999
Modified model (a.m./p.m. milk yield relation)	Fat	1	2 368	0.96	0.94	0.97	0.97
		2	967	0.95	0.93	0.96	0.96
		3	1 083	0.95	0.93	0.96	0.96
	Protein	1	2 368	0.998	0.998	0.998	0.999
		2	967	0.998	0.998	0.999	0.999
		3	1 083	0.999	0.999	0.999	0.999

Table 5 shows that protein is estimated more accurately than fat. The fat yield is estimated most accurately with the recommended model. But where it is not practically possible to register time, the a.m./p.m. milk yield relation can be used. It is almost impossible to see the difference on the accuracy of protein – whether it is the recommended model or the modified model that is used. It applies for both fat and protein that the yield is estimated most accurately if the milk samples are taken out alternate morning and evening.

Differences between the known and the estimated 305 days fat and protein yield are seen in table 6. The estimated yield was determined with the "AM-method" (Table 4) and the recommended model. If the difference is negative, the yield is overestimated and if the difference is positive, the yield is underestimated.

In table 6 it is shown that the differences for fat have a tendency to be negative, because the estimated 305 days fat yield on average was overestimated for all three parities. The differences for protein have a tendency to be positive, because 305 days protein yield was on average underestimated for all three parities.

Table 6. Differences (kg) per parity between known and estimated 305 days yield. The yields were estimated with the recommended model and the AM sampling method.

	Parity	Number yield	Mean known	Mean dif.	S.D.
305 days fat yield (kg)	1	2 368	317	-2.4	11.8
	2	967	359	-2.3	15.2
	3	1 083	376	-2.2	15.6
305 days protein yield (kg)	1	2 368	265	1.3	1.9
	2	967	298	1.0	2.0
	3	1 083	303	0.7	2.0

95 % of the estimates of the fat yield per day will be within the range of 150 grams from the true fat yield per day. 95 % of the estimations of the protein yield per day will be within the range of 35 grams from the true protein yield per day.

There is no sign that the model over- or underestimate certain groups – for example high-performance cows.

Conclusion

The recommended model describes a reasonable correlation between the true yield and the estimated yield for both fat and protein. The calculated accuracies (Acc) are just as high as the accuracies described in the literature. Higher accuracies are obtained using kg fat and kg protein in the model instead of fat and protein percentages.

It has been investigated whether a modification of the recommended model can give the same accuracy. The modification was that the time interval between the individual milkings were replaced by the relation between kg milk a.m. and kg milk p.m. The modified version of the model gives almost the same accuracy as the recommended model.

For both models the accuracy for kg fat was on the highest level with milk samples taken out in the evening. For protein, the models gave the best estimate based on samples taken in the morning. This indicated that the best results are obtained taking the samples out alternate morning and evening.

The analysis to determine the 305 days yield showed that the protein yield could be calculated with great accuracy based on one milk sample per test day. Lower accuracy existed on determination of the fat yield.

The model gave within this material a bit too high estimation of the fat yield in all lactations, but the protein yield was estimated a bit too low. The estimated yields were compared to the calculated true yield.

The breeding values for bulls are not affected from the changed sampling, but in the cows the bias in the estimation of the test day yield of the single cow will affect the breeding values. Thus the selection of bull dams has a bit lower accuracy.

Implementation

In Denmark it is decided to change the milk recording system in a way, where we use electronic Tru-Test milk meters, and per test day only one milk sample (morning or evening) is analyzed. Results are presented to the farmer as the measured yield in kg milk and the estimated percentages of fat and protein.

List of references

Schaeffer, L.R., Jamrozik, J., Van Dorp, R., Kelton, D.F. og Lazenby, D.W. 2000. Estimating daily yelds of cows from differnt milking schemes. Livestock Production Science, 65, pp. 219-227.