Case study 6: Field studies to evaluate the benefits on milk production of providing smallholder farmers in Kenya access to credit to allow shifting of concentrate feeding to early lactation

John Rowlands [1], Thomas Achia [2], Dannie Romney [1], Margaret Wambugu[3], Robert Kaitho[1,4], and David Njubi[1]

[1]International Livestock Research Institute, Nairobi, Kenya
[2]Department of Mathematics, University of Nairobi, Nairobi, Kenya
[3]Ministry of Agriculture and Rural Development, Nairobi, Kenya
[4]Kenya Agricultural Research Institute, Nairobi, Kenya

Contents

Summary

This case study describes a series of studies carried out with smallholder farmers involved in dairy production in the central highlands of Kenya. A previous survey in the region had reported low milk yields, probably due to low to inadequate feed supplies. It was decided to investigate whether this constraint could be removed by arranging for farmers to have access to credit from the cooperatives, to which they sold their milk, so that they could feed higher levels of concentrates to their cows in early lactation.

The studies began with an on-station trial to investigate an appropriate level at which concentrates could be reallocated so that all concentrates could be fed during the early part of lactation rather than throughout. This was followed by a cross-sectional questionnaire survey to determine current levels of milk yield, a 12-month participatory, longitudinal study in which some farmers were offered credit, others not, and finally a further questionnaire cross-sectional survey to study the impact of the intervention a year later. Concurrent with these studies were series of interviews with farmers to obtain information on how they were adopting the new feeding regime.

The case study explains how the design of the longitudinal study collapsed when some farmers from the control group started to follow what the others were doing. This resulted in

changes in the research process and caused complications in approaches to the statistical analysis.

On reading the study report (Romney et al., 2005) it was decided to question the adequacy of two of these approaches to the statistical analysis and to examine whether a more rigorous treatment of the data would lead to the same or a different interpretation. This is the primary focus of this case study.

Difficulties arose in the handling the large amounts of data that were collected. Data management and editing aspects arising from longitudinal studies are discussed. The programming attributes of GenStat are also illustrated in the statistical analysis.

Finally the importance of being able to critically review a scientific publication, and not to necessarily take the authors' statistical analyses and conclusions at their face value, is emphasised.

Background

A cross-sectional survey of 365 households (1.2% estimated sample) in Kiambu District in Central Province of Kenya (a primary source of milk for the Nairobi market) estimated that 77% of households were involved in dairy production (Staal et al., 1998). However, the survey also identified that milk yields were generally low. The farmers said that inadequate feed supplies were the major cause.

A 12-month longitudinal study that followed this survey confirmed farmer observations. Despite recommendations from agricultural extension officers that concentrates should be provided in early lactation and gradually increased until no further response in milk yield was observed, the study showed that farmers did not feed concentrates in this way.

In subsequent interviews with the farmers it became clear that the farmers were using feeds opportunistically and fed whatever was available. Fodder shortages, especially in the dry season, were usually replaced by concentrates, but farmers tended to purchase the cheapest forms of concentrate feed and not necessarily the most desirable ones

Typical quantities of just 2 kg/day of concentrates were fed at a flat rate throughout lactation with no extra feed given at the start of lactation when the need for nutrients was greatest. Reasons given for these low levels of concentrate offered were the high costs of concentrates and the difficulties in taking advantage of credit facilities to buy them.

With this background knowledge a series of studies was initiated to investigate whether milk production could be increased by changing the ways in which concentrates were fed. These studies are described in this Case study.

Rather than feed at a fixed rate throughout lactation it was proposed that it would be better for all the concentrates to be fed for a defined period during early lactation at a considerably higher rate than the 2 kg/day commonly used by farmers, and with none after this period had ended.

A field study was therefore planned to determine whether such a reallocation of concentrates was feasible under on-farm conditions and whether the dairy cooperatives to which farmers sold their milk would be willing to cooperate by providing concentrates on credit up front (Romney et al., 2005).

Research strategy

The paper by Romney et al. lays out the overall strategy for the research project. It was concluded that the following studies were needed:

1. An on-station experiment to plan and define the nature of the concentrate feeding intervention by quantifying the benefits of shifting all the feeding to a period during early lactation.
2. A cross-sectional, questionnaire survey to collect initial data on cow milk yields in smallholder farms.
3. A longitudinal field study with 'treatment' and 'control' farmers to evaluate the effect of reallocation of concentrates.
4. A follow-up, cross-sectional, questionnaire survey 12 months later to see how any improvements made during the previous year were being maintained.

The longitudinal study did not perform precisely as anticipated and further consultations were required with farmers to obtain additional retrospective information.

The first study was carried out on station to quantify the effect on milk yield of feeding concentrate to cows at 8 kg/day over the first 12 weeks of lactation with that of feeding 2 kg/day over 48 weeks, in other words a shift of the provision of the bulk of concentrates to early lactation, rather than giving it little at a time throughout lactation.

It was felt necessary to carry out this investigation before embarking on a field study.

Milk yields were found to be 45, 24, 6 and 8% lower between 1-75, 76-150, 151-225 and

226-305 days post calving, respectively, for cows on the flat rate method of feeding compared with those given all their concentrate feeding during the first 12 weeks.

With these positive findings a field study was planned to determine if reallocation of concentrates was feasible under on-farm conditions. The aim was to develop a participatory study that allowed the researchers to observe how farmers implemented the proposed strategy and to what extent farmers were prepared to follow the recommendations.

Early in 1999, 60 farmers were selected at random from a group of 90 volunteer farmers with cows expected to calve between April and July 1999. Another 30 who did not wish to test the scheme on their farms agreed to be monitored as a control group over the same period.

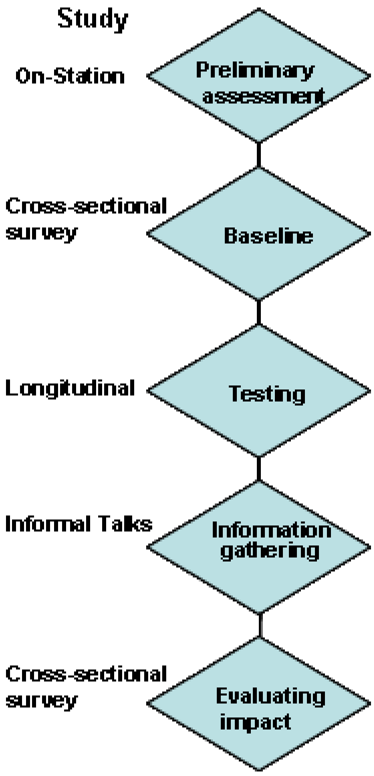
A questionnaire was prepared to provide baseline information including baseline data on milk yields of cows to be monitored in the survey.

Cows calving between March and October 1999 were then monitored over a period of at least 200 days post calving. Records were collected, generally at weekly intervals until the 12th week of lactation and fortnightly thereafter, on the quantities of concentrates and forage offered and the milk yield produced on that day.

A follow-up questionnaire was carried out 12 months after the end of the monitoring period. The purpose was to collect information on feeding practices and level of milk yield one year later.

Additional informal interviews were held with farmers during the longitudinal study to obtain information on the availability of credit, their approaches to concentrate feeding and additional information not being obtained through routine data collection.

Four studies of different types (on-station, cross-sectional questionnaire survey, longitudinal monitoring, and cross-section questionnaire survey) were thus conducted.



As already mentioned, informal talks were often held with farmers during the longitudinal monitoring visits to collect data, and this provided valuable additional information of a qualitative nature. An extra activity 'Information gathering' is thus included in the flow chart. The questionnaire developed for 'Evaluating impact' also provided an opportunity to collect retrospective data that had not been perceived to be necessary when the longitudinal study was planned.

The research process often works like this. A researcher can only see a certain distance ahead and must be prepared for plans to deviate from those anticipated at the outset.

Objectives

The primary objective is:

- To investigate the potential for shifting to early lactation the allocation of concentrates fed to cows by smallholder dairy farmers in the central highlands of Kenya by arranging access to credit from the dairy cooperatives to which they sell their milk.

A subsidiary objective necessary for an initial on-station study is:

- To determine the manner in which concentrate should be reallocated - namely the level that should be given and over what period the reallocation should take place.

Questions to be addressed

Results from the different studies are reported by Romney et al. (2005).

The data were difficult to analyse. For example, the longitudinal study was set up as a participatory study because it was considered important to understand the benefit of the concentrate feeding strategy when placed in the farmers' hands.

Following good biometric practice the researchers organised treatment and control groups to allow formal statistical inference. But the farmers did not do as expected.

Volunteer farmers did not follow the guidelines, and some of the control farmers, when seeing what the volunteer farmers were doing, did likewise. Thus the results of the study were driven by the actions of the farmers and the strategy for analysing the data had to be adjusted accordingly.

In view of the complexity of the data collected by Romney et al. (2005) we have decided to investigate the adequacy of two of the approaches to the statistical analysis and to examine whether a more rigorous treatment of the data would lead to the same or a different interpretation of the data.

Question1

Lactation curves were plotted in the paper for each of the three sets of data: baseline survey questionnaire, longitudinal survey and post-survey questionnaire.

Each set of data was analysed using Wood's model for a lactation curve:

$$y_n = an^b e^{-cn},$$

where y_n is the milk yield on day n and a, b and c are constants.

However, the structures of the data sets collected at the different times are different: one longitudinal and two cross-sectional. Wood's curve is strictly suitable only for repeated milk yields from groups of individual cows and is unable to handle the large variation in milk yield from cow to cow that cannot be distinguished within a set of single observations.

Therefore the first question that we shall address is:

- Are the curves produced in Figure 2 of Romney et al.'s paper appropriate representations of the patterns in the three years?

Question 2

As expected, there was a positive association between average milk yield and average concentrate fed during the trial (Figure 3 of Romney et al.). But this is an overall effect and doesn't take into account differences in patterns of feeding at different phases of lactation.

It might be worth examining the nature of the association at different stages of lactation to see whether any further light can be thrown on the influence of level of concentrate feeding on milk yield. We shall divide the lactation period into five 7-week phases.

The data from both volunteer and control farmers were pooled when it was realised that some control farmers were following the guidelines given to the volunteers. Nevertheless, there was considerable variation in the ways that farmers (both volunteer and control) followed the recommendations for concentrate feeding. It might be possible, therefore, to classify the data set into different levels of concentrate feeding in early and late lactation and to determine average lactation curves for different categories of farmers (low-low, high-low and high-high, say).

Therefore the second question we shall address is:

- Can more information be derived from the analysis of concentrate feeding and milk yield than that given in Romney et al.'s paper?

Study design

The design of the main longitudinal study is described in Romney et al. (2005). The concept of the study was presented in early 1999 to a meeting of farmers. Ninety farmers with cows due to calve between April and July 1999 volunteered to participate in the study and alter their feeding practices by reallocating their concentrate feeding to early lactation. Sixty from the 90 farmers were selected at random for the study. Another 30 farmers who did not wish to alter their feeding practices agreed to be monitored as controls.

The concentrate feeding management offered to the farmers was based on the earlier experimental findings, in other words to reallocate the concentrate they buy to the first 12 weeks of lactation.

However, there was no attempt to insist that the farmers followed exactly the instructions provided by the researchers. Instead, the researchers were interested in seeing how farmers put the instructions into practice. Frequency of monitoring was planned to be weekly for the first 12 weeks of lactation and fortnightly thereafter.

It is useful to consider here to what extent this trial comes under the category of participatory research. The majority of the data were collected by the research scientists, and not in a participatory manner. The researchers decided on the questions they wanted to ask and designed the questionnaires to collect the information they required. They analysed the data.

On the other hand, farmers were not prevented from making adjustments to the recommended concentrate feeding guidelines. Thus, the researchers had no control of how the feeding protocol was implemented. So, in this sense, the trial was participatory.

Indeed the level of participation that occurred resulted in essentially only one 'treatment' being applied without a control, when two (a treatment and a control) were planned and included in the design at the outset.

Thus, the idea of being able to analyse a study with the conventional, classical approach of comparing the means of a treatment and control were thwarted. Without the existence of a formal control, analytical tools needed to be devised to provide interpretations that looked at variability and took into account changes over time.

This will be a common feature of experiments where some participation by the farmer is expected. Researchers may need to be flexible in changing their approaches to data analysis from those envisaged when a study is first designed.

Source material

The raw data are contained in CS6Data1 with cow already defined as a factor. Three additional files have been created for use during the analysis. CS6Data2 is used for calculating the average relationship between milk yield and day of lactation during longitudinal monitoring. CS6Data3 is used for calculating associations between average milk yields and average level of concentrates fed over successive 7-week periods of lactation.

CS6Data4 contains milk yield data for each of the baseline, longitudinal monitoring and post- monitoring studies, and is used for comparing associations between milk yield and day of lactation across the three years.

Details of the contents of these files are stored in CS6Doc1, CS6Doc2, CS6Doc3 and CS6Doc4, respectively.

Data management

We do not go into detail here about the way the data were managed, except to report that difficulties were encountered in handling the vast amounts of quantitative data, in particular those arising during the longitudinal monitoring. To some extent this delayed analysis and the researchers had to rely on talks with farmers to gather information on the progress of the trial.

Where survey data are collected, especially of the longitudinal nature that featured in the main part of this study, it is essential that there are adequate human and computer resources to enter and process the data at the speed required by the project.

In retrospect, the researchers were perhaps somewhat ambitious in the quantities of data that they expected to be collected. Less intensive data collection would have resulted in an easier study to manage and could have achieved the same objectives.

Both the changes in the study design brought about by some control farmers deciding not to adhere to the feeding protocol, and the heavy data handling requirement that resulted in long gaps in getting reports from the analysis, made it necessary to obtain additional information from farmers and co-operatives

Additional data had to be collected retrospectively in a questionnaire prepared for the final cross-sectional survey to evaluate sustainability of the impact when it was realised that additional information on concentrate purchase and use was needed to understand some of the results that had been obtained.

The basic data that we shall use to address the two questions defined under Questions to be addressed are stored in CS6Data1.

We first need to check the raw data (in Raw data worksheet of CS6Data1) to make sure that all the data are suitable for the analyses required. It will be important in particular to make sure that there has been regular milk recording for each cow and that no long gaps have occurred, especially in early lactation.

When scanning columns of data by eye it is sometimes helpful to first unstack the data. Here we unstack the column for the variable day using **Spread → Manipulate → Unstack...**(see dialog box below), so that the values for each cow can be read in a different column. Immediately we see that there is a long gap, for instance, between days of lactation 115 and 197 for cow 24 (headed 'day_24' in the spreadsheet below).

Row	day_20	day_24	day_68	day_502	day_505	day_506	day_507	day_509	day_512	day_516	day_520	day_535
1	7	11	7	1	13	7	25	18	10	4	3	6
2	14	15	13	9	19	15	31	24	17	10	12	13
3	21	17	21	16	26	21	38	32	24	17	18	21
4	29	24	28	24	33	35	43	39	31	24	25	27
5	35	31	35	29	40	43	52	44	38	32	32	34
6	42	52	43	37	48	48	57	51	45	38	40	41
7	57	59	49	51	55	57	59	58	52	47	46	62
8	64	66	63	59	62	63	66	67	58	52	55	69
9	71	73	71	80	68	71	72	71	66	59	60	76
10	77	80	76	128	75	75	94	79	73	65	67	84
11	84	89	85	142	82	85	102	85	82	67	73	91
12	98	94	91	170	90	92	108	93	83	74	75	97
13	105	115	99	178	96	104	117	100	86	79	82	104
14	113	197	103	187	103	119	122	108	94	87	87	114

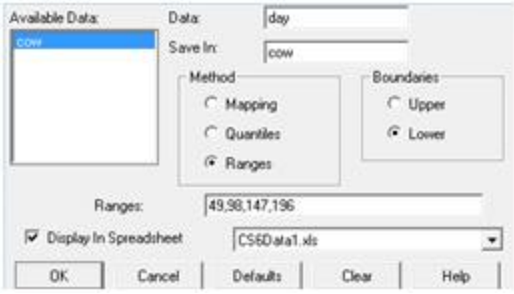
A few records, such as the record for day 197 for cow 24, have been removed (see Audit worksheet in CS6Data1). There was only one record for the two cows 549 and 807 during the first 7 weeks of lactation. As this period is crucial both to the study and to the analysis, we have decided to ignore these two cows.

An arbitrary decision was also made to remove records from a few cows with particularly long lactations. Measurements greater than 270 days of lactation have been ignored. An unusually high concentrate value of 16.1 kg fed to cow 801 on day 206 of lactation has also been deleted. This is probably a mistake in data entry.

When modifications are being made to a data set, many of an arbitrary nature, it is important, as done here, to make notes of the changes, so that someone else can see what was done and repeat the analysis in a different way, if necessary.

After reloading the 'edited' worksheet of CS6Data1 we find that GenStat has changed the dates into numerical formats. We can change the values back into dates using **Spread → Column →Attributes/Format…** and click 'cdate' and repeat for 'date'.

Next, we need to group the data into five lactation phases, four of 7 week intervals and a remainder, so that we can later analyse the association between milk yield and concentrate fed at different phases of lactation. We do this using **Data → Form Groups (Factors)...** and enter the values 49, 98, 147, 196 to define group endings (see dialog box below) This results in values of 24, 73, 120, 169, 225 representing the medians for the five groups. We can convert these into values 1, 2, 3, 4, 5 via **Spread → Factor → Change levels…**



We save this file as a GenStat worksheet in CS6Data2. This file provides the basis for fitting Wood's model in Statistical modelling to calculate the average relationship between milk yield and day of lactation.

However, the regression analysis needed to compare associations between mean milk yields and concentrates in individual phases will require mean values first to be calculated for each cow.

We shall prepare this data file CS6Data3 now. We start with **Stat → Summary Statistics → Summaries of Groups (Tabulations)…**, and put the variable milk into the 'Variate' box and cow and phase into the 'Groups' box. Then, by clicking the 'Save' button, storing the means in the variable milkmeans, and clicking the 'Display Tables in the Spreadsheet' box (see dialog box below), we can obtain a 2-way table of means displayed on the screen.

We need to convert the table to a vector format (**Spread → Manipulate → Convert…**) and rename the columns as milk1, milk2, milk3, milk4, milk5 as shown below.

We now repeat the same steps but with the variable conc instead of milk.

Having done this, and then reselected the first spreadsheet of means for milk yield, we can add the data for conc1to conc5 to the existing columns of means for milk by simply clicking **Spread → Add →Data in GenStat….**

To add the overall averages for milk yield and concentrate we need to repeat the above steps but with just the factor cow in the 'Groups' box. We then save the complete file having converted all fields to 2 decimal places in CS6Data3.

Row	!cow_1	milk1	milk2	milk3	milk4	milk5
1	20	9	8.5	7.75	6.25	6
2	24	16.6	14.7143	15	*	*
3	68	17.1667	21.125	19.5	16.2	10
4	502	15.3333	10.6667	5.5	10.3333	9
5	505	18.125	18.2143	14.2125	17.4	20.3667
6	506	25.0833	26.6667	24.25	22.125	21.5
7	507	17.875	18.3333	17.2143	14.375	13.25
8	509	12.3	12.2143	10.8333	9.5	8
9	512	13.6667	13.1875	10.9167	11.25	10.3125
10	516	13.9286	9.42857	6.42857	6.25	6.25
11	520	10.5	9.3125	5.7	4.8	4.75
12	535	9.75	9	8.625	8.5	*
13	538	10.9	12.375	9.14286	8.375	7.33333
14	541	10.5714	10.2143	8.75	8.25	6.83333
15	550	10.4286	7.07143	6.33333	*	*
16	601	12.5	10.8571	9.33333	8.9	7.66667
17	604	8.8	10	9.16667	7	6.2

Statistical modelling

Contents

Fitting Wood's curve

Here we repeat the analysis conducted in Romney et al.'s paper by fitting Wood's model: $y_n = an^b e^{-cn}$, where y_n is the milk yield on day n and a, b and c are constants. This model has been successfully applied to milk yields of dairy cows raised in the developed world, and satisfactorily represents the peak yield that is commonly achieved around about six to seven weeks into lactation.

In sub-Saharan Africa, where smallholder farmers often cannot afford the optimal feeding levels required in early lactation, the early peak in milk yield often does not materialise, and so Wood's formula may not always provide such a good fit (Kamidi, 2005).

It is common to convert Wood's model to logarithms:
$\log_e y_n = \log_e a + b. \log_e n - c.n$

which can be fitted by a multiple regression analysis with independent variables
$x_1 = \log_e n$ and $x_2 = n$.

Romney et al. (2005) also include terms in the model to account for seasonal effects of month of calving and month of sampling but, for the purpose of this case study, we shall not consider these effects.

First we calculate the logarithmic values in CS6Data2 using Spread → Calculate → Column… In doing so we discover that the first milk yield recorded for cow 830 is zero. Therefore, we need to delete this record and make a note in the Audit worksheet of CS6Data1 (see note).

We need to analyse each cow separately. The best way to do this is to unstack the columns for logmilk, logday and day (**Spread → Manipulate → Unstack…**) but this time we shall tick the box 'Create Suffixed column names' in the dialog box.

The section of the spreadsheet below displays the first few records for logmilk for the first nine cows. The remainder of the logmilk values are to the right off the screen, followed by values for logday and day also hidden.

Row	logmilk_1[1]	logmilk_1[2]	logmilk_1[3]	logmilk_1[4]	logmilk_1[5]	logmilk_1[6]	logmilk_1[7]	logmilk_1[8]	logmilk_1[9]	logmilk
1	2.20	2.89	2.53	2.08	2.72	3.07	2.77	2.64	2.71	2.
2	2.30	2.67	2.83	2.89	2.89	3.18	2.83	2.56	2.56	2.
3	2.08	2.89	2.89	2.86	2.89	3.35	2.97	2.30	2.53	2.
4	2.20	2.80	2.86	2.86	2.83	3.20	2.94	2.40	2.67	2.
5	2.20	2.77	2.80	2.83	2.97	3.24	2.92	2.53	2.64	2.
6	2.20	2.64	3.07	2.64	3.04	3.28	2.67	2.48	2.56	2.
7	2.40	2.64	3.16	2.64	3.02	3.28	3.09	2.56	2.74	2.
8	2.14	2.64	3.09	2.20	2.86	3.26	2.97	2.48	2.56	2.
9	1.95	2.74	3.06	2.20	2.94	3.30	3.00	2.53	2.48	2.
10	1.95	2.67	3.00	1.79	2.89	3.24	2.74	2.48	2.71	2.
11	2.20	2.71	3.00	1.61	2.92	3.37	2.83	2.53	2.53	2.
12	2.08	2.77	3.00	2.30	2.83	3.26	2.80	2.44	2.40	2.
13	1.87	2.71	2.97	2.30	2.83	3.30	2.94	2.44	2.53	2.
14	1.95	*	2.89	2.40	2.71	3.22	2.80	2.40	2.64	2.

Having set the new spread sheet as the active worksheet we can now do a multiple regression analysis for the first cow by applying Stats → Regression Analysis → Linear Models… . We can also use the 'Save' button to store values for the three estimates in the vector Est[1]. The analysis is shown below.

***** Regression Analysis *****
Response variate: logmilk_1[1]
Fitted terms: Constant, logday_1[1], day_1[1]
*** Summary of analysis ***

	d.f.	s.s.	m.s.	v.r.
Regression	2	0.6661	0.33307	18.18
Residual	23	0.4215	0.01832	
Total	25	1.0876	0.04350	

Percentage variance accounted for 57.9
Standard error of observations is estimated to be 0.135
*** Estimates of parameters ***

	estimate	s.e.	t(23)	t pr.
Constant	2.162	0.227	9.51	<.001
logday_1[1]	0.0342	0.0712	0.48	0.636
day_1[1]	-0.002681	0.000915	-2.93	0.008

	d.f.	s.s.	m.s.	v.r.
Regression	2	0.6661	0.33307	18.18
Residual	23	0.4215	0.01832	
Total	25	1.0876	0.04350	

Percentage variance accounted for 57.9
Standard error of observations is estimated to be 0.135
*** Estimates of parameters ***

	estimate	s.e.	t(23)	t pr.
Constant	2.162	0.227	9.51	<.001
logday_1[1]	0.0342	0.0712	0.48	0.636
day_1[1]	-0.002681	0.000915	-2.93	0.008

If we now look at the GenStat Input file we can see how the instructions are represented

```
MODEL logmilk_1[1]
TERMS [FACT=9] logday_1[1],day_1[1]
FIT [PRINT=model, summary, estimates; CONSTANT=estimate;\
FPROB=yes; TPROB=yes; FACT=9] logday_1[1],day_1[1]
RKEEP ; ESTIMATES=Est[1]
```

We copy these statements into an Input window and modify them by incorporating a loop 'For [NTIMES=77]' and replacing the [1] in each of the variate names by [I]. As we shall need to calculate average values for the three estimates we can incorporate statements to do this with the final values stored in the variate 'MeanEst'.

```
VARIATE [NVALUES=3;VALUES=0,0,0] SumEst,MeanEst
SCALAR [VALUE=0] I
FOR [NTIMES=77]
CALCULATE I=I+1
MODEL logmilk_1[I]
TERMS [FACT=9] logday_1[I],day_1[I]
FIT [PRINT=model, summary, estimates; CONSTANT=estimate;\
FPROB=yes; TPROB=yes; FACT=9] logday_1[I],day_1[I]
RKEEP ; ESTIMATES=Est[I]
CALCULATE SumEst = SumEst + Est[I]
ENDFOR
CALCULATE MeanEst = SumEst/77
PRINT MeanEst
```

We then use Run →Submit Selection to complete the calculations.

The PRINT statement at the end of the series of statements shows that the average values for the estimates are:

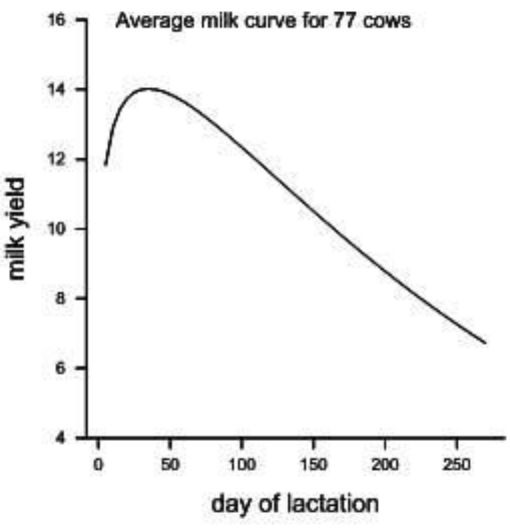
MeanEst
2.2427
0.1562
-0.0045

We can now calculate average fitted values by preparing the statements below. We define a variable nday to have values between 5 and 270 at 5 day intervals and calculate the predicted value for average milk yield (avemilk) for each value of nday., ,

```
VARIATE [NVALUES=54; VALUES=(1...54)] nday,avemilk
CALCULATE nday = nday*5
CALCULATE avemilk = MeanEst$[1] + MeanEst$[2]*LOG(nday)+MeanEst$[3]*nday
CALCULATE avemilk = EXP(avemilk)
```

Note the facility in GenStat for using the $ syntax for identifying individual levels of a vector.

Finally we can use **Graphics → Lineplot…** (selecting avemilk and nday from the drop-down lists) to produce the graph shown. We see that the average curve has the shape expected with a peak for milk yield at about 6 weeks. The maximum is a little higher than the curve shown in Figure 2 of Romney et al. but this may be due to a slightly different approach to the analysis. We have also not taken into account effects of month of calving and sampling on milk yield.



As already indicated, the fitting of Wood's model to the single set of milk values obtained pre and post the longitudinal survey is not really appropriate. It is better to fit a simple linear regression and compare slopes and intercepts.

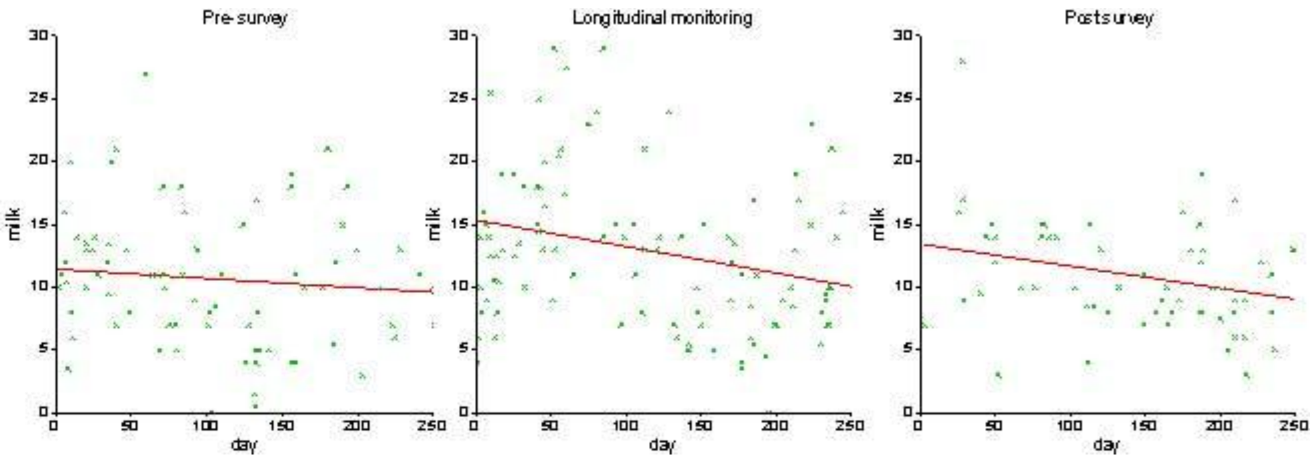
We have put the milk values for the pre- and post-surveys into separate worksheets in CS6Data4, having restricted the range of days post calving to between 1 and 270.

Calving dates for the monitored cows ranged from 3rd March to 8th October, 1999 - not for the full year. Thus, in any one month there was not a complete range of days across the lactation period.

In order to produce a similar data set to the other two years, it was decided to use the first milk yields recorded in June and December, respectively, for the cows that were sampled in those months, and then to pool these two data sets. These data have also been stored in CS6Data4.

Milk curve comparisons

By fitting linear regressions in GenStat by **Stats → Regression Analysis → Linear Models…** and clicking the 'Options…' button, followed by the 'Plot Fitted Model', the graphs below can be produced. The lengths of the x- and y- axes have been made the same for each graph to allow direct comparisons between them. This is important - otherwise the different scales determined by the different ranges in data values can often give a false impression



The three intercepts and slopes are summarised below. It can be seen that there were average reductions in milk yield of 0.21 (P<0.01) and 0.18 (P<0.5) kg/day during lactation in the second and third years but there was no significant reduction in the first year. The higher intercepts in the second pair of years also indicate how higher milk yields were being obtained in early lactation in these years

These results confirm the findings in Figure 2 of Romney et al. However, taking into consideration the limitations in the data, they probably provide a more reasonable representation of the patterns that can be presented.

<u>Pre-survey</u>

	estimate	s.e.	t(77)	t pr.
Constant	11.43	1.06	10.82	<.001
day	-0.00741	0.00797	-0.93	0.355

<u>Longitudinal monitoring</u>

	estimate	s.e.	t(102)	t pr.
Constant	15.34	1.04	14.70	<.001
day	-0.02116	0.00726	-2.91	0.004

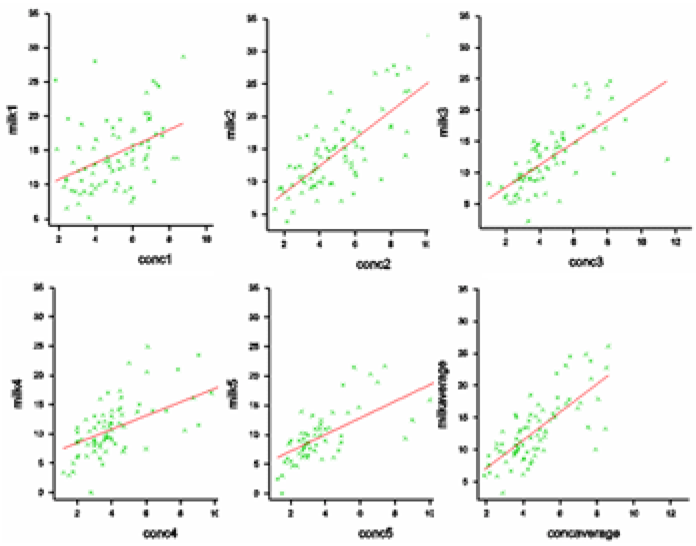
<u>Post-survey</u>

	estimate	s.e.	t(63)	t pr.
Constant	13.43	1.18	11.43	<.001
day	-0.01759	0.00703	-2.50	0.015

Concentrates and milk yield

Here we use the data that have been arranged in CS6Data3 (as described under Data management) and fit linear regressions for each of the five phases of lactation and for the five phases combined. Once again the scales for each of the graphs have been modified so that the same scale is used for each graph.

Apart from a few outliers in the bottom right hand corners of the graphs for phases 4 and 5 of lactation (x-axis: 'conc4' and conc5'), there is generally a good association between concentrate fed and milk yield produced. The poorest association is possibly during the first seven weeks of lactation ('conc1'), namely before peak yield is reached.



A summary of the regression coefficients for each phase of lactation, together with the correlation coefficients, is shown in the table below. Each correlation coefficient was calculated by taking the square root of the 'Percentage variance accounted for' divided by 100 in the linear regression output of GenStat.

The table shows that the strongest association was in Phase 2 of lactation (r=0.73).

The full regression equation for the average values was found to be:

Milkaverage = 2.75 (± 1.15) + 2.18 (± 0.23) concaverage

This is very similar to the equation shown in Figure 3 of **Romney et al**. The regression coefficient for Phase 2 was similar to the average regression coefficient across the whole of lactation. Thereafter, the regression coefficient decreased in value.

Dependent variable	Regression coefficients	Correlation coefficient
milk1	1.21 ± 0.31 conc1	0.39
milk2	2.10 ± 0.22 conc2	0.73
milk3	1.78 ± 0.21 conc3	0.69
milk4	1.16 ± 0.21 conc4	0.54
milk5	1.41 ± 0.21 conc5	0.64
avemilk	2.18 ± 0.23 concaverage	0.73

We also carried out a number of multiple regression analyses Stats → Regression analysis → Generalized linear models … with 'conc2' combined with 'conc3', 'conc4' and 'conc5' in turn. By comparing the regression coefficients for 'conc2' with the coefficients for the other variables it can be seen that the concentration fed in Phase 2 between weeks 7 and 14 had as strong an influence on subsequent milk yields as the concentrate fed at the time.

Dependent variable	Regression coefficients	Correlation coefficient
milk3	1.78 ± 0.21 conc3	0.69
milk3	1.03 ± 0.26 conc3 +1.06 ± 0.25 conc2	0.76
milk4	1.16 ± 0.21 conc4	0.54
milk4	0.58 ± 0.21 conc4 + 1.17 ± 0.23 conc2	0.69
milk5	1.41 ± 0.21 conc5	0.64
milk5	0.98 ± 0.25 conc5 + 0.83 ± 0.27 conc2	0.70

Romney et al. reported that cooperative prices for concentrates varied from 8 to 15 Kenya shillings(Ksh)/kg and for milk from 14.5 to 18 Ksh/kg. They noted that with a regression coefficient of 2 kg/day farmers would make a profit, even at the highest prices for concentrates and the lowest prices for milk.

The present analysis confirms this to be the case during early to mid lactation but suggests that profit margins might have not have been so great later in lactation. Indeed, some farmers may have been over-feeding their cows at this time.

Average levels of concentrations fed to cows decreased only slightly during lactation - from about 5 kg/day in early lactation to slightly under 4 kg/day in the latter part of lactation. Thus, farmers on the whole did not follow the researchers' recommendations and continued to feed concentrates throughout lactation.

			Average concentrate fed (kg/day)			
Phase	Weeks	No. cows	Mean	Median	Minimum	Maximum
1	1-7	77	5.05	5.07	1.84	8.76
2	8-14	77	4.93	4.60	1.50	10.16
3	15-21	76	4.60	4.00	1.00	11.52
4	22-28	73	4.28	3.70	1.20	12.21
5	Remainder	58	3.84	3.18	1.23	10.40
Average		77	4.70	4.39	1.90	8.62

There were large variations in concentrate feeding among farmers, perhaps reflecting the mixture of control and volunteer farmers in the group.

It would also appear that not all control farmers followed the guidelines given to volunteer farmers and continued to feed low levels of concentrates to their cows at any stage of lactation.

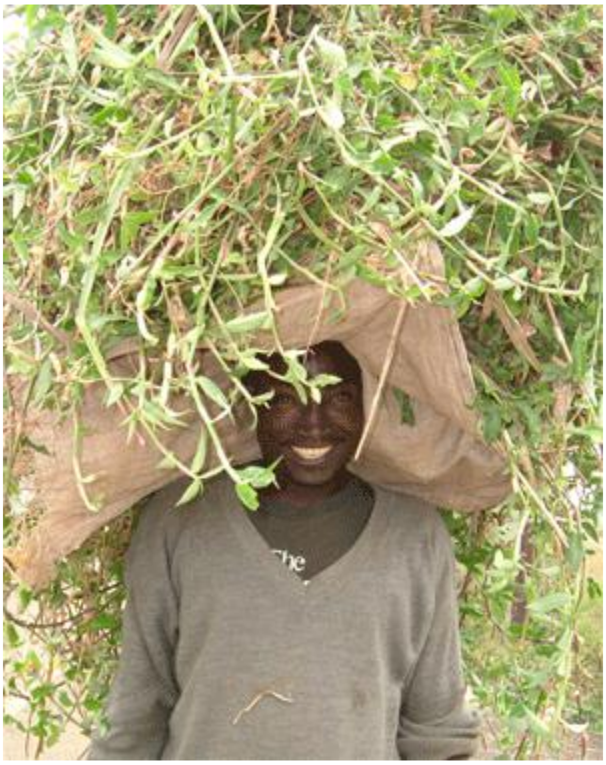
As suggested earlier in this case study, it might be interesting to categorise farmers into groups according to the average levels of concentrates fed (say, high-high, high-low, low-low for early-late lactation) and fit a lactation curve to each group. One might expect the shape of the curve to be flatter for cows fed low concentrates throughout than for those fed higher concentrations at the start.

This is addressed by one of the study questions.

Findings, implications and lessons learned

- The findings from the statistical analysis comparing the shapes of the lactation curves between the three periods essentially supported the conclusions reported by Romney et al., but our methods were somewhat sounder and provided more reasonable representations of the data.
- The statistical analysis of the differences in patterns of associations between milk yield and concentrate provided more information on the ways that patterns varied with stage of lactation, and gave some insight into the degrees to which farmers were adhering to the recommended feeding strategy.
- As a reader of a scientific publication it is imperative that one is able to review critically the contents of the paper and not accept at face value the authors' conclusions. It is particularly important that a reader is able to question the validity of any statistical methods that the authors may have applied.
- A strategy for tackling a research project is not cast in stone. The researcher needs to be ready for unexpected findings, unexpected delays, disappointments when studies do not work out as planned, and so on. When farmers decided not to follow precisely the recommendations given in this project, the researchers had to revisit their strategy for analysing the data.
- Furthermore, the realisation that additional data were required that had not been planned meant that additional information based on farmer recall (less reliable than from current observations) had to be collected by formal questionnaire at the time of the final cross-sectional survey

- Participatory studies inevitably lead to complexities in data analysis and interpretation. Whilst such studies get closer to the real life situation this needs to be borne in mind when designing such a study.
- One should never underestimate the data handling requirements that result from longitudinal monitoring studies. Well designed data management systems need to be planned, and sufficient human resources to handle the data input assured.



Study questions

- Using CS6Data2 categorise cows into high-high; low-high; high-low and low-low concentrate-fed groups. The high-high etc. periods should correspond to weeks 1-14 and weeks 15 onwards, respectively. Having calculated mean values for each cow over these two periods you will need to decide on suitable high/low ranges for each period so that there are reasonable numbers of cows in each group. Justify your groupings.

- Using the results of question 1, calculate average Wood lactation curves for each group. Comment on any differences in regression coefficients. How might you determine whether the coefficients are statistically significant?
- Prepare a letter that you might send to the authors of the Romney et al. publication explaining your observations from the analyses carried out in this case study.
- Prepare a report suitable for presenting to the farmers outlining some of the results obtained in this case study.
- Describe the checks that you would include in a database system that you might have written to handle the data collected during the monitoring period. Describe any interim summary reports that you think would be useful to have during the course of the study that might have been informative to the researchers.
- Read the paper by Romney et al. again. Under 'Conclusions: Methodological critique:' they state "The analytical tools used allowed us to interpret the information collected despite there being no formal control. Rather than control for underlying variability, the study attempted instead to record the variability and use this to explain results." Discuss what you think the authors mean. Do you agree with the authors' statement?
- Write an essay on different forms of farmer participation in agricultural research. You may use examples described in Case Study 1 and Case Study 10.
- Review the paper by Kamidi (2005) and write a referee's report.

Related reading

Kamidi, R.G. 2005. Biometry of lactation trends in dairy cattle. Paper presented at the *9th Bi-Annual Conference of The Sub-Saharan African Network (SUSAN) of the International Biometrics Society,* Addis Ababa, Ethiopia, 2005. 8 pp. Full text

Romney, D., Wambugu, M., Kaitho, R., Biwott, J., Chege, L., Omore, A., Staal, S., Wanjohi, P., Njubi, D. and Thorpe, W. (2005) Case study D: Improving the efficacy of concentrate usage by smallholder dairy farmers in Kenya In. *Participatory Livestock Research A Guide,* Conroy, C (ed) ITDG Publishing Rugby, pp185-195. Full text

Staal, S., Chege, L., Kenyanjui, M., Kimari, A., Lukuyu, B., Njubi, D., Owango, M., Tanner, J., Thorpe, W. and Wambugu, M. 1998. *Characterisation of dairy systems supplying the Nairobi milk market: a pilot survey in Kiambu District for the identification of target groups of producers.* Smallholder Dairy (R&D) Project Report. International Livestock Research Institute, Nairobi, Kenya., 55 pp. Full text

Acknowledgements

The research reviewed within this case study was carried out through the Smallholder Dairy (R&D) Project (SDP) of the Kenya Ministry of Agriculture and Rural Development, the Kenya Agricultural Research Institute (KARI) and the International Livestock Research Institute. SDP was funded by the UK Department for International Development (DFID) and worked closely with many others, including farmers, dairy co-operatives, agricultural extension agents, ministry regulatory officers.

We are grateful to Romney et al. for permitting their paper to be discussed in this case study.