Calculation of reference values for herd health indicators

Aurélien Madouasse

Wednesday, September 24, 2014

# Work outline

The aim of this document is to introduce herd health indicators and to show how they can be calculated from datasets that are commonly available on farm.

The sources of data that are used are animal registration and movement data, milk recording data and artificial insemination data.

The indicators that are going to be presented below are:

* Productive life indicators
* Mortality indicators
* Reproductive indicators
* Mastitis indicators
* Indicators linked to diet

# How to contribute ?

This file was generated using a reporting program with which R code can be embedded within written text. I this document, the R code I wrote appears as shaded and explanations appear as normal text. To use this facility you need to have the following programs installed on your computer: R, Rstudio and the R package Knitr. You also need to have the 4 data files copied into a data directory nested in your R working directory. Once all this is in place, it is fairly easy to use, provided you understand R code. If you have all this, you can replicate this analysis and explore other indicators or play with your own data. If anyone is interested, I am happy to move further in that direction.

A simpler way to work is to use the track change function on the attached MS Word file.

We also attach a file named 'indicators.R' which contains all the R code that is shown in the current document.

# Provisional schedule

## Data description

Task: describe available data

Scheduled: 26/09/2014

Completed: 26/09/2014

## Productive life indicators

Task: describe procedure for calculating indicators

The indicator is:

* Length of productive life

Scheduled: 26/09/2014

Completed: 17/10/2014

## Mortality indicators

Task: describe procedure for calculating indicators

The indicators are:

* Mortality rate for calves
* Mortality rate for young stock
* Mortality rate for cows

Scheduled: 26/09/2014

Completed: 17/10/2014

## Reproductive indicators

Task: describe procedure for calculating indicators

The indicators are:

* Age at first calving
* Calving interval

Scheduled: 26/09/2014

Completed: 26/09/2014

## Mastitis indicators

Task: describe procedure for calculating indicators

The indicators are:

* Test-day proportion of somatic cell counts above threshold
* Between test-days incidence of raised somatic cell counts

Scheduled: 26/09/2014

Completed: 26/09/2014

## Indicators linked to diet

Task: describe procedure for calculating indicators

The indicators are:

* Test-day proportion of fat to protein ratios above theshold
* Test-day proportion of fat to protein ratios below theshold

Scheduled: 26/09/2014

Completed: 26/09/2014

# Data

In order for me to create and for you to replicate the indicator variables from herd data, I have extracted data from 10 randomly selected herds. For these herds,the aim was to gather all the data that were relevant for performing the analyses between 2003 and end of 2010. The data are in 4 separate variables that I am going to present below.

All the dates are formatted as yyyy-mm-dd.

## Cow dataset

Data on individual cows are stored in a variable called *anim*.

head(anim)

## cowID date\_birth breed  
## 1 1 2000-09-19 56  
## 2 2 2000-04-03 56  
## 3 3 1999-01-11 66  
## 4 4 1999-01-24 56  
## 5 5 1999-03-05 66  
## 6 6 1999-08-29 66

The columns are:

* *cowID*: unique cow identifier.
* *date\_birth*: date of birth
* *breed*: cow breed

## Milk recording dataset

The milk recording dataset is stored in the *rec* variable.

head(rec)

## herdID date\_mr cowID date\_calv parity milk bfat prot scc  
## 1 1 2003-01-24 3 2002-12-18 3 296 429 298 156  
## 2 1 2003-01-24 4 2002-09-09 2 160 593 371 821  
## 3 1 2003-01-24 5 2002-11-10 2 238 534 305 60  
## 4 1 2003-01-24 6 2002-09-13 1 274 488 284 120  
## 5 1 2003-01-24 7 2002-08-31 1 282 453 319 92  
## 6 1 2003-01-24 8 2002-09-04 1 174 617 306 102

The columns are:

* *herdID*: unique herd identifier
* *date\_mr*: date of milk recoring
* *cowID*: unique cow identifier
* *date\_calv*: calving date
* *parity*: cow parity
* *milk*: milk yield in kg
* *bfat*: butterfat content in g/kg
* *prot*: protein content in g/kg
* *scc*: somatic cell count in 100 cells/mL

## Artificial inseminations dataset

Artificial insemination data are stored in the *ais* variable.

head(ais)

## herdID cowID date\_pCALV date\_AI  
## 1 1 1633 0000-00-00 2000-05-09  
## 2 1 1619 2000-03-03 2000-05-12  
## 3 1 1629 0000-00-00 2000-05-17  
## 4 1 1633 0000-00-00 2000-06-30  
## 5 1 1629 0000-00-00 2000-07-15  
## 6 1 1633 0000-00-00 2000-07-22

The columns are:

* *herdID*: unique herd identifier
* *cowID*: unique cow identifier
* *date\_pcalv*: date of previous calving
* *date\_AI*: date of artificial insemination

## Movement data

Animal movement data are stored in the *mvt* variable.

head(mvt)

## herdID cowID date\_IN cause\_IN date\_OUT cause\_OUT  
## 1 1 1 2001-08-14 A 2007-05-29 B  
## 2 1 2 2001-08-14 A 2006-08-07 B  
## 3 1 3 1999-01-11 N 2004-12-06 B  
## 4 1 4 1999-01-24 N 2004-07-19 B  
## 5 1 5 1999-03-05 N 2004-04-20 B  
## 6 1 6 1999-08-29 N 2005-02-10 B

The columns are:

* *herdID*: unique herd identifier
* *cowID*: unique cow identifier
* *date\_IN*: date cow moved into this herd
* *cause\_IN*: reason for cow moving into this herd
* *date\_OUT*: date cow moved out of this herd
* *cause\_OUT*: reason for cow moving out of this herd

Here is a list of the codes used in the *cause\_IN* variable:

|  |  |
| --- | --- |
| code | Meaning |
| A | Purchase |
| N | Birth |
| P | Lending |

In our dataset, the frequencies of each reason for joining a herd are as follows:

table(mvt$cause\_IN)

##   
## A N P   
## 279 2442 3

Here is a list of the codes used in the *cause\_OUT* variable:

|  |  |
| --- | --- |
| code | Meaning |
| B | Sold for meat |
| C | Slaughtered for consumption of farm |
| E | Sold to other farm |
| M | Dead |

In our dataset, the frequencies of each reason for leaving a herd are as follows:

table(mvt$cause\_OUT)

##   
## B C E M   
## 564 1285 6 670 199

# Periodicity of indicator calculation

For the calculation of indicators we specify a starting date and an end date. These are stored in the start\_date and end\_date variables respectively.

start\_date <- "2006-01-01"  
end\_date <- "2006-12-31"

# Productive life indicators

## Length of productive life

The aim of the productive life indicator is to measure the time a cow spends on the herd between her first calving and her culling or death.

The difficulty with this indicator is that we need to know when the cow first calved. This piece of data can be unavailablable at the time of calculation. For example, a cow that is ten years old today can have had herd first calving 8 years ago.

One way to go can be to use the parity of culled cows. For example the median parity of the cows culled between start\_date and end\_date.

# Mortality indicators

## Mortality of calves

With this indicator, we want to measure the proportion of calves which die between birth and 30 days (mort\_calves\_age1) of life. For this indicator, we only take female calves into account because male calves can be sold at various ages, which are rarely known. Including these male calves would make the calculations very imprecise.

We could either work with calf-days at risk or with the frequency of deaths in calves that were present during the investigated time interval, whether it was at least one day or for the entire time interval.

The difference between these 2 options becomes less important as the period investigated grows bigger. That is to say it is important if the calculations are carried out monthly but much less so if they are done on an anual basis.

### Denominator

### Numerator

## Mortality of young stock

Here we are interested in the proportion of deaths in calves between mort\_calves\_age1 and the first calving.

### Denominator

### Numerator

## Mortality of cows

Here we are interested in the number of cows who died on farm as a proportion of all the cows present on the farm during the year. This last figure needs to be estimated.

## Cows present on farm

Cows are present on farm from their first calving until culling, death or sale.

### Denominator

### Numerator

# Reproductive indicators

## Creation of a list of lactations

Most of the calculations undertaken in this part are easier if we have a list of individual lactations containing cow IDs, dates of calving and parity. This is done as follows.

We use the milk recording dataset as it contains all the necessary information.

lac <- rec[, c("herdID", "cowID", "date\_calv", "parity")]  
lac <- lac[-which(duplicated(lac)),]

## Age at first calving

Animals included are all the nulliparous cows that calved between start\_date and end\_date. For these animals, we calculate the median age at calving.

We need to pieces of information:

* calving dates of the nulliparous cows that calved between start\_date and end\_date from the lac dataset.
* birth dates of the above identified cows from the anim dataset.

From the lactation dataset, the cows that calved as nulliparous are now cows of parity 1.

lac\_p1 <- lac[lac$parity == 1, -match("parity", colnames(lac))]

The dates of birth for these cows are incorporated from the anim dataset.

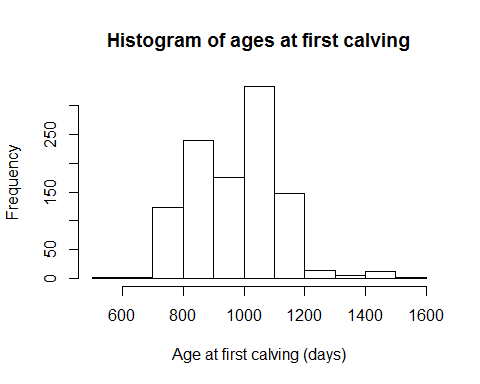
lac\_p1 <- merge(lac\_p1, anim[, c("cowID", "date\_birth")], by = "cowID", all.x = TRUE)

The ages at first calving can now be calculated.

lac\_p1$age\_calv1 <- with(lac\_p1, as.numeric(as.Date(date\_calv) - as.Date(date\_birth)))

The distribution of the ages at first calving in this dataset is presented below.

hist(lac\_p1$age\_calv1,  
 main = "Histogram of ages at first calving",  
 xlab = "Age at first calving (days)")



The percentiles 25, 50 and 75 of the distribution of ages at first calving can be calculated as follows.

quantile(lac\_p1$age\_calv1, c(.25, .5, .75), na.rm = TRUE)

## 25% 50% 75%   
## 851 994 1083

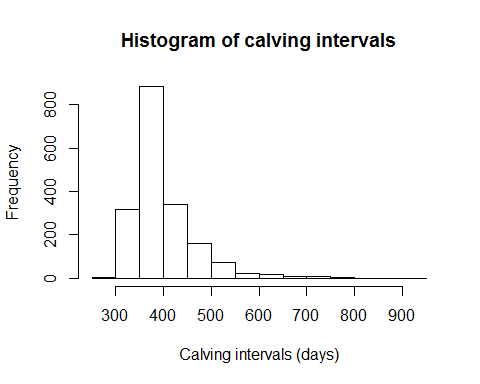
## Calving interval

The calving intervals are calculated from the dataset containing the list of lactations, lac. As a first step, we add the date of the previous calving that we call date\_pcalv. The number of days between the latest calving and the prevous calving is then calculated.

lac <- lac[order(lac$cowID, as.Date(lac$date\_calv)),]  
lac$pcowID <- c(NA, lac$cowID[-nrow(lac)])  
lac$date\_pcalv <- c(NA, as.character(lac$date\_calv)[-nrow(lac)])  
lac$calv\_int <- with(lac, ifelse(cowID == pcowID,  
 as.numeric(as.Date(date\_calv) - as.Date(date\_pcalv)),  
 NA))  
lac <- lac[, -match(c("pcowID", "date\_pcalv"), colnames(lac))]  
lac <- lac[order(lac$herdID, as.Date(lac$date\_calv)),]

The distribution of calving intervals is presented below.

hist(lac$calv\_int,  
 main = "Histogram of calving intervals",  
 xlab = "Calving intervals (days)")



The percentiles 25, 50 and 75 of the distribution of calving intervals can be calculated as follows.

quantile(lac$calv\_int, c(.25, .5, .75), na.rm = TRUE)

## 25% 50% 75%   
## 357 380 422

# Mastitis indicators

Our primary and possibly only source of data concerning mastitis will be somatic cell count from milk recording. These somatic cell counts are measured at the cow level, on a monthly basis, most of the time. The procedures to collect and analyse milk recording data, and consequently somatic cell counts, are described by the international committee for animal recording (ICAR).

Various thresholds are used to categorise cows as having mastitis. A common threshold is 200 000 cells/mL following work undertaken by Dohoo in the 1990s. But in France, farmers are used to a threshold of 300 000 cells/mL because it is what is used by the national milk recording organisation. Here, I will use a threshold of 200 000 cells/mL. I set it explicitely in the code below so that you can modify it if wanted.

scc\_T <- 200

A problem that has not been considered here is: how to take cows that were missing on a test-day into account. They could have missing data because they were treated for mastitis.

## Test-day proportion of somatic cell counts above threshold

Once we have milk recording data and a threshold, it is extremely easy to determine the proportion of milk recordings above the theshold.

* For the whole data:

with(rec, length(scc[scc > scc\_T]) / length(scc))

## [1] 0.2674

* Between start\_date and end\_date:

with(rec[as.Date(rec$date\_mr) >= start\_date & as.Date(rec$date\_mr) <= end\_date,],   
 length(scc[scc > scc\_T]) / length(scc))

## [1] 0.2742

* Between start\_date and end\_date, by herd:

with(rec[as.Date(rec$date\_mr) >= start\_date & as.Date(rec$date\_mr) <= end\_date,],   
 tapply(scc, herdID, function(x) length(x[x > scc\_T]) / length(x)))

## 1 2 3 4 5 6 8   
## 0.4561 0.1262 0.4297 0.2722 0.1166 0.3695 0.1069

## Between test-days incidence of raised somatic cell counts during lactation

There could be various ways to calculate this indicator. Raised SCC has to be defined using SCC collected at an earlier point in time. We can use the previous milk recording of a cow. In this case, dry periods will be included. Therefore, we will define this indicator by comparing SCCs on 2 consecutive test-days, provided days in milk on the second test-day is greater than days in milk on the first test-day.

First we create variables for days in milk at milk recording (Dim), days in milk at previous test-day (pDim) and somatic cell count at previous test-day (pscc).

rec <- rec[order(rec$cowID, as.Date(rec$date\_mr)),]  
rec$Dim <- with(rec, as.numeric(as.Date(date\_mr) - as.Date(date\_calv)))  
rec$pcowID <- c(NA, rec$cowID[-nrow(rec)])  
rec$pdate\_mr <- c(NA, as.character(rec$date\_mr)[-nrow(rec)])  
rec$pdate\_mr <- with(rec, ifelse(cowID == pcowID, pdate\_mr, NA))  
rec$pDim <- c(NA, rec$Dim[-nrow(rec)])  
rec$pDim <- with(rec, ifelse(cowID == pcowID, pDim, NA))  
rec$pscc <- c(NA, rec$scc[-nrow(rec)])  
rec$pscc <- with(rec, ifelse(cowID == pcowID, pscc, NA))  
  
rec <- rec[, -match(c("pcowID", "pdate\_mr"), colnames(rec))]

We can now determine the number of cows moving from below to above the threshold between consecutive test-days, as a proportion of the number of cows that were initially below the threshold.

* For the whole dataset:

with(rec[!is.na(rec$scc) & !is.na(rec$pscc) & rec$Dim > rec$pDim,],  
 length(scc[scc >= scc\_T & pscc < scc\_T]) / length(scc[pscc < scc\_T]))

## [1] 0.1156

* Between start\_date and end\_date:

with(rec[as.Date(rec$date\_mr) >= start\_date & as.Date(rec$date\_mr) <= end\_date & !is.na(rec$scc) & !is.na(rec$pscc) & rec$Dim > rec$pDim,],  
 length(scc[scc >= scc\_T & pscc < scc\_T]) / length(scc[pscc < scc\_T]))

## [1] 0.1008

* Between start\_date and end\_date, by herd: Here we create an indicator variable that is defined for cows that had a somatic cell count below the threshold at the previous test-day.

rec$Irais <- rep(NA)  
  
rec$Irais[!is.na(rec$scc) & !is.na(rec$pscc)] <- with(rec[!is.na(rec$scc) & !is.na(rec$pscc),], ifelse(pscc < 200, 0, Irais))  
  
rec$Irais[!is.na(rec$Irais)] <- with(rec[!is.na(rec$Irais),],  
 ifelse(scc >= 200, 1, 0))  
  
with(rec[!is.na(rec$Irais) & rec$Dim > rec$pDim,],   
 tapply(Irais, herdID, function(x) length(x[x == 1]) / length(x)))

## 1 2 3 4 5 6 7 8   
## 0.20464 0.07529 0.11314 0.10771 0.10643 0.15531 0.17313 0.10169

## Incidence of raised somatic cell counts during the dry period

Now that we have an indicator of raised somatic cell count during lactation, it is easy to apply it to determine the incidence of raised somatic cell count during the dry period. We just have to put as a condition that Dim < pDim.

with(rec[!is.na(rec$Irais) & rec$Dim < rec$pDim,],   
 tapply(Irais, herdID, function(x) length(x[x == 1]) / length(x)))

## 1 2 3 4 5 6 7 8   
## 0.25862 0.04846 0.15116 0.17045 0.13450 0.20000 0.35294 0.11602

## Percentage of cure during the dry period

We define cure during the dry period by somatic cell count going from above to below the threshold between the last milk recording in a lactation and the first milk recording in the following lactation.

rec$Icur <- rep(NA)  
  
rec$Icur[!is.na(rec$scc) & !is.na(rec$pscc)] <- with(rec[!is.na(rec$scc) & !is.na(rec$pscc),], ifelse(pscc >= 200, 0, Icur))  
  
rec$Icur[!is.na(rec$Icur)] <- with(rec[!is.na(rec$Icur),],  
 ifelse(scc < 200, 1, 0))  
  
with(rec[!is.na(rec$Icur) & rec$Dim < rec$pDim,],   
 tapply(Icur, herdID, function(x) length(x[x == 1]) / length(x)))

## 1 2 3 4 5 6 7 8   
## 0.5111 0.9014 0.4088 0.8110 0.7901 0.7500 0.9000 0.6867

# Indicators linked to diet

## Creation of the fat to protein ratio variables

rec$fpr <- with(rec, bfat / prot)

## Test-day fat to protein ratio as a measure of subclinical ketosis

The fat to protein ratio can be used as an indicator of ketosis. Two things need to considered to use this indicator : a threshold that defines ketosis (fpr\_ket\_T) and a time period after calving when ketosis is more likely (between fpr\_ket\_lbDim and fpr\_ket\_ubDim).

fpr\_ket\_T <- 1.4  
fpr\_ket\_lbDim <- 30  
fpr\_ket\_ubDim <- 100

* For the whole dataset:

with(rec[!is.na(rec$fpr) & rec$Dim >= fpr\_ket\_lbDim & rec$Dim <= fpr\_ket\_ubDim,],   
 length(fpr[fpr > fpr\_ket\_T]) / length(fpr))

## [1] 0.2399

* Between start\_date and end\_date:

with(rec[!is.na(rec$fpr) & rec$Dim >= fpr\_ket\_lbDim & rec$Dim <= fpr\_ket\_ubDim,],   
 length(fpr[fpr > fpr\_ket\_T]) / length(fpr))

## [1] 0.2399

* Between start\_date and end\_date, by herd:

with(rec[!is.na(rec$fpr) & rec$Dim >= fpr\_ket\_lbDim & rec$Dim <= fpr\_ket\_ubDim,],   
 tapply(fpr, herdID, function(x) length(x[x > fpr\_ket\_T]) / length(x)))

## 1 2 3 4 5 6 7 8   
## 0.32305 0.18994 0.28363 0.18302 0.06659 0.37094 0.05344 0.38023

## Test-day fat to protein ratio as a measure of subacute ruminal acidosis

The fat to protein ratio can also be used as a marker of subacute ruminal acidosis. In this case, a different threshold is used. The number of cows below the threshold is calculated.

fpr\_sara\_T <- 1.0

* For the whole dataset:

with(rec[!is.na(rec$fpr) ,],   
 length(fpr[fpr < fpr\_sara\_T]) / length(fpr))

## [1] 0.07468

* Between start\_date and end\_date:

with(rec[!is.na(rec$fpr) & as.Date(rec$date\_mr) >= start\_date & as.Date(rec$date\_mr) <= end\_date, ],   
 length(fpr[fpr < fpr\_sara\_T]) / length(fpr))

## [1] 0.05722

* Between start\_date and end\_date, by herd:

with(rec[!is.na(rec$fpr) & as.Date(rec$date\_mr) >= start\_date & as.Date(rec$date\_mr) <= end\_date, ],   
 tapply(fpr, herdID, function(x) length(x[x < fpr\_sara\_T]) / length(x)))

## 1 2 3 4 5 6 8   
## 0.02198 0.03095 0.11217 0.01036 0.16035 0.03741 0.05791