# 31

# DESCRIPTION OF DATASETS

All datasets used in the examples and sample problems in this text are provided for pedagogical purposes only. They are provided so that the reader can recreate the examples included in the text. Contributors have made data available to the readers of this text on this understanding and consequently, this is the only use for which they are provided.

In some cases, datasets have been modified since the initial publication of results from the study which generated the data. In many cases, only a subset of the original data (*ie* a subset of variables or a subset of observations) are included. Consequently, the reader should not expect to be able to duplicate results obtained in the original publication.

Some of the datasets described below do not feature in examples used in this edition of the book, but have either been used in the 1st edition or have been found to be very useful in courses covering material presented in this text.

In the descriptions that follow, unless otherwise specified, all variables coded 0 or 1 (0/1) have the following meaning:

0 = no, absent or negative

1 = yes, present or positive

All datasets can be downloaded from the Veterinary Epidemiologic Research website (http://www.upei.ca/ver).

The authors extend their sincere thanks and appreciation to the contributors of these datasets.

## ap2

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Håkan Vigre | single cohort | 1114 | pig |

### Reference(s)

Vigre H, Dohoo IR, Stryhn H, Busch ME. Intra-unit correlations in seroconversion to *Actinobacillus pleuropneumoniae* and *Mycoplasma hyopneumoniae* at different levels in Danish multisite pig production facilities Prev Vet Med. 2004; 63: 9-28.

### Brief description

Data were collected on 1,114 pigs from 35 batches produced on six farms that employed an 'all-in, all-out' production process. Pigs were weighed and blood sampled at the time of transfer from the weaner barn to the finisher barn (approximately 70 days of age) and again 6 weeks later (shortly before slaughter). Blood samples were tested for antibodies to *Actinobacillus pleuropneumonia* (Type 2), *Mycoplasma hyopnueumonia*, the influenza virus and the porcine respiratory and reproductive syndrome virus (PRRS). Two of the objectives of the study were to determine when seroconversion to the various agents occurred and at which level of the population (*eg* pig, batch or herd) most of the variation in seroconversion occurred.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| farm_id | farm identification | |
| batch_id | batch identification number | |
| litt_id | litter identification number | |
| pig_id | pig identification | |
| parity | farrowing number of sow | |
| vacc_mp | batch vaccinated against M. hyopneumoniae | 0/1 |
| seas_fin | season pigs in finishing unit (0=summer, 1=winter) | 0/1 |
| age_t | pig age at transfer from weaning to finishing unit | days |
| w_age_t | weight at age_t | kg |
| age_t6 | age plus approx. 6 weeks | days |
| w_age_t6 | weight at age_t6 | kg |
| dwg_fin | daily weight gain between age_t and age_t6 | gm |
| ap2_t | serological reac. against A. pleuropneumoniae serotype 2 at age_t | 0/1 |
| mp_t | serological reac. against M. hyopneumoniae at age_t | 0/1 |
| infl_t | serological reac. against influenza virus at age_t | 0/1 |
| prrs_t | serological reac. against PRRS virus at age_t | 0/1 |
| ap2_t6 | serological reac. against A. pleuropneumoniae serotype 2 at t6 | 0/1 |
| mp_t6 | serological reac. against M. hyopneumoniae at age_t6 | 0/1 |
| infl_t6 | serological reac. against influenza virus at age_t6 | 0/1 |
| prrs_t6 | serological reac. against PRRS virus at age 16 | 0/1 |
| ap2_sc | seroconversion to ap2 during the finishing period | 0/1 |

## beef_ultra

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Greg Keefe | single cohort | 487 | animal |

**Reference(s)**

Keefe G, Dohoo I, Valcour J, Milton R. Ultrasonic imaging of marbling at feedlot entry as a predictor of carcass quality grade Canadian Journal of Animal Science. 2004; 84: 165-70.

### Brief description

Data were collected on 487 cattle at the time that they entered a feedlot for ‘fattening’ prior to slaughter. Data consisted of demographic information plus readings obtained from an ultrasonic evaluation of the animal. Ultrasound measurements of backfat thickness, loineye area and the percentage of intramuscular fat (‘marbling’) were obtained. The objective of the study was to determine if ultrasound examination of the animal at the time of entry into a feedlot was able to predict final carcass grade (AAA, AA or A). Carcass grade depends primarily on the amount of intramuscular fat in the carcass at the time of slaughter.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| farm | farm id | |
| id | animal id | |
| grade | carcass grade | 1 = AAA<br>2 = AA<br>3 = A |
| breed | breed (known or estimated) | multiple |
| sex | gender | 0 = female<br>1 = male |
| bckgrnd | animal backgrounded | 0/1 |
| implant | hormone implant used | 0/1 |
| backfat | backfat thickness | mm |
| ribeye | area of rib eye muscle | sq cm |
| imfat | intramuscular fat score | % of area |
| days | fattening period | days |
| carc_wt | carcass weight | kg |

## bst_mast, bst_milk

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Ian Dohoo | meta-analysis | 29 , 28 | group of cows |

**Reference(s)**

Dohoo IR, DesCôteaux L, Leslie K, Fredeen A, Shewfelt W, Preston A, Dowling P. A meta-analysis review of the effects of recombinant bovine somatotropin. 2. Effects on animal health, reproductive performance, and culling Can J Vet Res. 2003a; 67: 252-64.

Dohoo IR, Leslie K, DesCôteaux L, Fredeen A, Dowling P, Preston A, Shewfelt W. A meta-analysis review of the effects of recombinant bovine somatotropin. 1. Methodology and effects on production Can J Vet Res. 2003b; 67: 241-51.

### Brief description

On request from Health Canada, the Canadian Veterinary Medical Association established an expert panel to review the production and health effect of recombinant bovine somatotropin (rBST) in dairy cattle. The panel carried out a meta-analysis of all available literature and evaluated a wide range of production and health effects. The data in this file consist of risk ratios for clinical mastitis that were associated with the use of rBST. Data from 29 distinct groups of cows, from 20 separate studies are included. The precision of the point estimate is included in the form of 95% confidence limits.

### Table of variables - bst_mast

| Variable | Description | Codes/units |
|---|---|---|
| study | study number | |
| group | cow group number | |
| parity | parity group | 1 = primiparous<br>2 = all ages combined<br>3 = multiparous |
| study_yr | year of study | |
| rr | risk ratio | |
| cilow | lower 95% confidence limit | |
| cihigh | upper 95% confidence limit | |
| dur | duration of treatment | days |
| dose_day | daily dosage | mg/day |

### Table of variables – bst_milk (same as above except for....)

| Variable | Description | Codes/units |
|---|---|---|
| diff | difference in milk production | |
| se | standard error of difference | |
| ncows | number of cows in study | |

## bvd_test

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Ann Lindberg | single cohort | 2,162 | cow |

**Reference(s)**

Lindberg A, Groenendaal H, Alenius S, Emanuelson U. Validation of a test for dams carrying foetuses persistently infected with bovine viral-diarrhoea virus based on determination of antibody levels in late pregnancy Prev Vet Med. 2001; 51: 199-214.

**Brief description**

Blood or milk samples were collected from 2,162 pregnant cows at various stages of lactation. Following the birth of their calf, the status of the calf with regard to persistent infection (PI) with the bovine viral diarrhea (BVD) virus was determined. The blood and milk samples were tested using an ELISA to determine levels of BVD virus antibodies in the cow. A variety of cutpoints were then examined to determine which gave the best combination of sensitivity and specificity for detecting PI+ calves. Logistic regression was used to evaluate the effects of other factors (particularly stage of lactation) on the estimated sensitivity and specificity of the test.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| cow_id | cow identification | |
| breed | breed | 1 = red and white<br>2 = black and white<br>3 = beef<br>4 = other |
| parity | parity group | 1 = primiparous<br>2 = all ages combined<br>3 = multiparous |
| pregmon | pregnancy month at test | |
| season | calving season | 1 = winter<br>2 = spring<br>3 = summer<br>4 = autumn |
| spec | type of specimen | 0 = milk<br>1 = blood |
| calfst | calf status | 0/1 |
| od | optical density | |
| co_5 | test result dichotomised at 0.5 | 0/1 |
| co_6 | test result dichotomised at 0.6 | 0/1 |
| | . . . etc . . . | |
| co_15 | test result dichotomised at 1.5 | 0/1 |
| co_16 | test result dichotomised at 1.6 | 0/1 |

## calf

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Jeanne Lofstedt | retrospective cohort | 254 | calf |

### Reference(s)

Lofstedt J, Dohoo IR, Duizer G. Model to predict septicemia in diarrheic calves J Vet Intern Med. 1999; 13: 81-8.

### Brief description

These data come from a retrospective analysis of the medical records from all diarrheic calves which were presented to Atlantic Veterinary College, PEI, Canada between 1989 and 1993. The ultimate objective of the study was to develop a logistic model which would predict whether or not the calf was septic at the time of admission (septic calves have a much poorer prognosis than non-septic calves and are not usually worth treating, given economic considerations).

There are 254 observations (records) and 14 variables in the dataset (calf). The original dataset had far more variables (including a lot of laboratory data) but this dataset contains only a subset of the demographic data and the physical examination data collected. All observations were determined on the day of admission, except for the outcome (sepsis) which was based on all data available at the time of death or discharge.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| case | hospital case number | |
| age | age at admission | days |
| breed | breed | coded 1-9 |
| sex | sex | 0 = female 1 = male |
| attd | attitude of calf | 0 = bright, alert<br>1 = depressed<br>2 = unresponsive, comatose |
| dehy | % dehydration | |
| eye | uveitis/hypopyon clinically evident | 0/1 |
| jnts | swollen joints clinically evident | number of joints affected |
| post | posture of calf | 0 = standing<br>1 = sternal<br>2 = lateral |
| pulse | pulse rate | beats per minute |
| resp | respiratory rate | breaths per minute |
| temp | rectal temperature | oC |
| umb | swollen umbilicus clinically evident | 0/1 |
| sepsis | sepsis (outcome) | 0/1 |

## calf_pneu

| **Contributor(s)** | **Study type** | **# records** | **Unit of record** |
|---|---|---|---|
| Iver Thysen | cohort | 24 | calf |

### Reference(s)

Thysen I. Application of event time analysis to replacement, health and reproduction data in dairy cattle research Preventive Veterinary Medicine. 1988; 5: 239-50.

### Brief description

These published data were used in one of the early publications in the veterinary literature discussing the use of survival analysis techniques. The data consist of mortality records from 24 calves that were housed in one of two housing systems: continuous housing, or batch (*ie* all-in all-out) housing. The event of interest was the time to onset of clinical pneumonia. Calves were followed for up to 150 days.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| calf | calf id | |
| stock | stocking method | 0 = batch<br>1 = continuous |
| days | time to onset of pneumonia or censoring | days |
| pn | pneumonia | 0/1 |

## clin_mast

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Richard Olde Riekerink | longitudinal | 5338 | case of mastitis or censoring event |

**Reference(s)**

Olde Riekerink RG, Barkema HW, Kelton DF, Scholl DT. Incidence rate of clinical mastitis on Canadian dairy farms JDairy Sci. 2008; 91: 1366-77.

**Brief description**

Data were collected from 106 herds from 10 Canadian provinces over a period of approximately 1½ years. Producers collected milk samples from all cases of clinical mastitis and these samples were frozen and then collected (approximately monthly) and shipped to the Atlantic Veterinary college, UPEI for culturing. The outcome of interest for the subset of data used in this text was the time from calving (or previous case of mastitis) to each clinical case. Only cows with at least 100 days of follow-up have been included. A total of 743 cases were observed in 608 of the 4595 cows that were monitored during the study.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| cowid | cow number | |
| prov | province | categorical - 1-10 |
| herd | herd number | |
| pr | parity (lactation number) | 1 to 14 |
| start | start period at risk | date |
| stop | end period at risk | date |
| par | period at risk | total time at risk |
| cases | number of mastitis cases | |
| d | days from calving to case or censoring | days |
| mast | mastitis case or censoring | 0 = cesnored<br>1 = clinical mastitis |

## culling

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Ashwani Tiwari | longitudinal | 721 | cow |

### Reference(s)

Tiwari A, VanLeeuwen JA, Dohoo IR, Stryhn H, Keefe GP, Haddad JP. Effects of seropositivity for bovine leukemia virus, bovine viral diarrhoea virus, Mycobacterium avium subspecies paratuberculosis, and Neospora caninum on culling in dairy cattle in four Canadian provinces Vet Microbiol. 2005; 109: 147-58.

### Brief description

These data were collected as part of a prevalence survey of 4 infectious agents in dairy herds in 5 Canadian provinces, but only data from one province are included in this dataset. Only the test results for *Mycobacterium avium* subspecies paratuberculosis (MAP or Johne's disease) is included in the dataset used in this text. Serum samples were collected from 30 cows within each herd (if available) and tested for MAP. Cows were followed for 4 years following testing and whether or not the cow was culled in that time period was recorded. Cows that were sold for dairy purposes were excluded from the analysis. Ultimately, records from 721 cows from 30 herds were available for analysis.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| herd | farm identification number | |
| cow | cow identification number | |
| dar | days after testing | days |
| lact_c3 | lactation 1, 2, or 3+ | |
| johnes | seropositive for Johne's | 0/1 |
| culled | culled within 3.5 yr. follow-up | 0 = censored<br>1 = culled |

## dairy_ids

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| John VanLeeuwen | survey (cross-sectional) | 2454 | cow |

**Reference(s)**

VanLeeuwen J, Keefe G, Tremblay R, Power C, Wichtel J. Seroprevalence of infection with Mycobacterium avium subspecies paratuberculosis, bovine leukemia virus and bovine viral diarrhea virus in Maritime Canada dairy cattle Canadian Veterinary Journal. 2001; 42: 193-8.

**Brief description**

These data were collected as part of a prevalence survey of four infectious diseases of dairy cattle in eastern Canada. 30 herds in each of three provinces (Prince Edward Island, Nova Scotia and New Brunswick) were randomly selected from lists of all dairy herds participating on a milk-production monitoring program. Within each herd, approximately 30 animals were randomly selected and blood samples collected. These samples were tested for antibodies to: *Neospora caninum*, *Mycobacterium avium (*subsp. *paratuberculosis)* and enzootic bovine leukosis virus. In addition, a group of non-vaccinated heifers were bled and tested for bovine virus diarrhea virus, but these test results are not included in this dataset. Sampling weights were computed as the inverse of the product of the probability of a herd being selected and the probability of a cow being selected within a herd.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| prov | province | |
| herd | herd identification number | |
| cow | cow identification number | |
| lact | lactation number | |
| dim | days in milk | days |
| johnes | Johne's test result | 0/1 |
| leukosis | leukosis test result | 0/1 |
| neospora | neospora test results | 0/1 |
| tot_hrd | total herds in province | |
| prob_hrd | probability of herd being selected | |
| tot_cow | total cows in herd | |
| tot_smpl | total cows sampled in herd | |
| prob_cow | probability of cow being selected | |
| prob_smp | overall probability of a cow being selected | |
| weight | sampling weight | |

## daisy2

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| John Morton | single cohort (longitudinal) | 9383 | lactation |

**Reference(s)**

Several in preparation at time of writing

**Brief description**

These data are a subset of the data collected during a large prospective longitudinal study carried out in 9 regions of Australia. Within each region, a dairy practitioner was responsible for herd selection and overseeing the study. Both year-round and seasonal calving dairy herds were included in the study. All lactations starting in the follow-up period of 12-15 months were enrolled. Disease and insemination data were recorded by producers with data entry taking place at the veterinary clinics. Milk production data were collected electronically from milk recording agencies. The data in this subset consisted of 9,383 lactation records from 8,441 cows in 42 year-round calving herds from 4 geographic regions. For many of the analyses in this text, a subset of 7 herds with high rates of reproductive diseases was used.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| region | Region | georaphic region (coded 1-4) |
| herd | Herd Number | |
| cow | Cow number (unique) | |
| study_lact | Study lactation number | 1st or 2nd lactation in the study period |
| herd_size | Herd size | |
| mwp | Minimum wait period for herd | days |
| parity | Lactation number | |
| milk120 | Milk volume (l) in first 120 days of lactation | litres |
| calv_dt | Calving date | |
| cf | Calving to first service interval | days |
| fs | Conception at first service | 0/1 |
| cc | Calving to conception interval | days |
| wpc | Interval from wait period to conception | days |
| spc | Services to conception | |
| twin | Twins born | 0/1 |
| dyst | Dystocia at calving | 0/1 |
| rp | Retained placenta at calving | 0/1 |
| vag_disch | Vaginal discharge observed | 0/1 |
| h7 | Indicator for 7 herd subset | |

### elisa_repeat

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Javier Sanchez | experimental | 40 | milk sample |

**Reference(s)**

Sanchez J, Dohoo IR, Markham F, Leslie K, Conboy G. Evaluation of the repeatability of a crude adult indirect Ostertagia ostertagi ELISA and methods of expressing test results Vet Parasitol. 2002; 109: 75-90.

**Brief description**

Forty individual cow milk samples were repeatedly tested (six times) using a crude *Ostertagia* antigen indirect ELISA. Results were recorded both as raw optical density (OD) values and values adjusted based on the readings for the positive and negative controls in the plate.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| id | sample identification | |
| raw1 | raw OD - sample #1 | |
| raw2 | raw OD - sample #2 | |
| | . . .etc . . . | |
| raw6 | raw OD - sample #6 | |
| adj1 | adjusted OD - sample #1 | |
| adj2 | adjusted OD - sample #2 | |
| | . . .etc . . . | |
| adj6 | adjusted OD - sample #6 | |

### fec

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Ane Nødtvedt | single cohort | 2,250 | monthly fecal egg count |

**Reference(s)**

Nødtvedt A, Dohoo I, Sanchez J, Conboy G, DesCôteaux L, Keefe G, Leslie K, Campbell J. The use of negative binomial modelling in a longitudinal study of gastrointestinal parasite burdens in Canadian dairy cows Can J Vet Res. 2002; 66: 249-57.

**Brief description**

Monthly (in some herds less frequently) fecal egg samples were collected from lactating age dairy cows (*n*=313) in 38 herds over a period of 1 year. The data were collected as part of a multifaceted study into parasitism in lactating dairy cows which included a longitudinal epidemiologic investigation and a controlled trial of the effects of deworming at calving with eprinomectin. The effects of factors at the sampling-day, cow and herd levels on fecal egg counts were evaluated.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| province | Canadian province | 1 = PEI<br>2 = Quebec<br>3 = Ontario<br>4 = Saskatchewan |
| herd | herd identifier | |
| cow | unique cow identifier | |
| visit | visit number | |
| tx | eprinomectin treatment at calving | 0/1 |
| fec | fecal egg count | eggs/5 gm |
| lact | lactation | 0 = primiparous<br>1 = multiparous |
| season | season | 1 = oct-dec 99<br>2 = jan-mar 00<br>3 = apr-jun 00<br>4 = jul-sep 00 |
| past_lact | lactating cows have access to pasture | 0/1 |
| man_heif | manure spread on heifer pasture | 0/1 |
| man_lact | manure spread on cow pasture | 0/1 |

## feedlot

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Wayne Martin | case-control | 588 | animal |

### Reference(s)

Martin SW, Nagy E, Armstrong D, Rosendal S. The associations of viral and mycoplasmal antibody titres with respiratory disease and weight gain in feedlot calves Can Vet J. 1999; 40: 560-7, 570.

### Brief description

This dataset represents the combined data from a number of studies on the role of specific micro-organisms as causes of bovine respiratory disease (BRD). Typically these beef cattle enter feedlots in the fall of the year and approximately 30% will develop BRD. The general strategy for the studies was to bleed all of the animals on arrival at the feedlot and again 28 days later (since most of the occurrence of BRD occurs in that time period). For analyses, we used all of the samples from cases and an approximately equal number from controls. In some of the smaller groups we used all samples and hence in these the study design was essentially a single cohort. The studies were conducted at essentially the same feedlots in different years, but depending on their size, one feedlot could have numerous groups of calves on the study in any given year. The titres were recorded in a quantitative manner but have been dichotomised in this dataset. **Note** At the time these data were collected, one of the important bacteria was called Pasteurella hemolytica. It is now referred to by its newer name Mannheima hemolytica.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| group | group identification | |
| tag | eartag number | |
| province | province of feedlot | 1 = Alberta<br>2 = Ontario |
| brd | clinical bovine respiratory disease (case-control) | 0/1 |
| brsvpos | arrival titre to brsv | 0/1 |
| brsvsc | seroconversion to brsv during study | 0/1 |
| bvdpos | arrival titre to bvd virus | 0/1 |
| bvdsc | seroconversion to bvd during study | 0/1 |
| ibrpos | arrival titre to ibr virus | 0/1 |
| ibrsc | seroconversion to ibr virus during study | 0/1 |
| pipos | arrival titre to pi3 virus | 0/1 |
| pisc | seroconversion to pi3 virus during study | 0/1 |
| phcypos | arrival titre to Mh cytotoxin | 0/1 |
| phcysc | seroconversion to Mh cytotoxin during study | 0/1 |
| phaggpos | arrival titre to Mh agglutinins | 0/1 |
| phaggsc | seroconversion to Mh during study | 0/1 |
| hspos | arrival titre to Hs | 0/1 |
| hssc | seroconversion to Hs during study | 0/1 |
| wt0 | arrival weight | kg |
| wt28 | 28-day weight | kg |

## fish_trial

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Tim Burnley | randomised controlled trial | 2000 | fish |

**Reference(s)**

Burnley T: Atlantic salmon vaccine performance and production characteristics evaluated through a multisite clinical field trial. *Unpublished PhD thesis.* Univ of PEI, Charlottetown, PEI Canada; 2009.

**Brief description**

These data are a small subset of a dataset collected in aquaculture clinical field trials in the Bay of Fundy, Canada, during the years 2004-2007. The objective of the research was to compare the performance of different vaccines on the growth and survival of Atlantic salmon under standard production conditions. The data here originate from a single cage of salmon that were individually tagged (by passive integrated transponder tags) in February 2005 and followed though to harvest in August 2007. Vaccines were randomly allocated to the fish and applied at day 0; thus, the different vaccine groups were held within the same cage throughout the production. Weight measurements were obtained on the fish at four sampling events (including the initial tagging) and at harvest. The data here consists of 100 fish from each of four vaccine groups, selected among fish with a full set of weight measurements to include all fish with jaw deformity and a random sample of the fish with no jaw deformity. As jaw deformities are formed at an early stage of growth and are not healed, detection of a jaw deformity at any sampling event (typically one of the earlier events) implied that the fish was labelled as having a jaw deformity throughout the entire growth period. The outcome of interest is the weight; as weights increase from means of 60 g at day 0 to 5,800 g at harvest, the analysis will be carried out on natural log scale.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| fish | fish id | |
| sample | sampling # | 1-5 |
| day | days since vaccination | |
| wt | sampling day weight | g |
| lnwt | natural log of weight | |
| wt_gain | weight gain since last sampling | g |
| adg | average daily gain | g/day |
| vaccine | vaccine group | 1-4 |
| jaw | jaw deformity | 0/1 |

## isa_day, isa_wk

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Larry Hammell, Ian Dohoo | retrospective cohort | isa_day 690<br>isa_wk 101 | cage - day<br>cage - week |

### Reference(s)

Hammell KL, Dohoo IR. Mortality patterns in infectious salmon anaemia virus outbreaks in New Brunswick, Canada J Fish Dis. 2005a; 28: 639-50.

Hammell KL, Dohoo IR. Risk factors associated with mortalities attributed to infectious salmon anaemia virus in New Brunswick, Canada J Fish Dis. 2005b; 28: 651-61.

### Brief description

Data for these studies were collected on the 1996 smolt year-class from 218 net-pens (sea-cages) from 14 sites, but only data from 9 cages at 1 site are included in this dataset. Mortality data for the spring – fall period of 1997 were recorded. The number of mortalities was recorded at each time they were collected from the cage by a diver. The total number of mortalities was then attributed equally on a daily basis over the interval since that last dive. Spikeplots of mortalities were generated and start, peak and end periods of outbreaks were subjectively identified from the observed pattern. Only data for the period of the outbreak are included in the dataset.

### Table of variables—isa_day

| Variable | Description | Codes/units |
|---|---|---|
| site | Site ID | |
| cage | Cage ID | |
| int_st_dt | Interval start date | date |
| int_end_dt | Interval end date | date |
| mortint | Interval length | days |
| morts | # of mortalities during interval | |
| par | Population at Risk | |
| mort_d | Daily mortality | |
| dt | Date (day during the interval) | |
| stage | Stage of outbreak | 1 – pre-peak<br>2 = post-peak |

The daily mortality data were subsequently compiled on a weekly basis and the outbreak pattern of each ISA outbreak was reconstructed based on the assumption that fish that died from ISA had been infectious for the previous 2 weeks and exposed the week before that. These data are in the file **isa_wk.**

### Table of variables—isa_week

| Variable | Description | Codes/units |
|---|---|---|
| site | Site | |
| cage | Cage | |
| week | Week of outbreak | |
| stage | Stage of outbreak | 1 – pre-peak<br>2 = post-peak |
| par | Population at risk | |
| C | # of cases (deaths) | |
| S | # susceptible | |
| E | # exposed | |
| Ia | # infectious - week one | |
| Ib | # infectious - week two | |
| I | # infectious (Ia + Ib) | |
| R | # removed | |
| N | # in population | |

## isa_lcm

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Pascale Nerette | longitudinal | 403 | fish |

**Reference(s)**

Nerette P, Dohoo I, Hammell L. Estimation of specificity and sensitivity of three diagnostic tests for infectious salmon anaemia virus in the absence of a gold standard. J Fish Dis. 2005; 28: 89-99.

**Brief description**

Replicate tissue samples from 403 fish from 4 different populations of Atlantic salmon that were assumed to have very different prevalences of ISA were collected. Samples were distributed to various laboratories in a blind fashion so that the labs did not know which population the fish came from. Samples were tested using IFAT (1 lab), PCR (3 labs) and virus isolation ( 2 labs) and the results used to build latent class models to evaluate the performance of all tests.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| site | Site identification | |
| cage | Cage identification | |
| id | Unique fish identification | |
| pop | Population (prevalence) | |
| ifat | IFAT score | 0-4 |
| ifat_c2 | IFAT 0/1 | 0/1 |
| pcr_a | PCR from lab A | 0/1 |
| pcr_b | PCR from lab B | 0/1 |
| pcr_c | PCR from lab C | 0/1 |
| vi_a | VI from lab A | 0/1 |
| vi_b | VI from lab B | 0/1 |

## isa_risk

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Larry Hammell, Ian Dohoo | cross-sectional | 182 | sea cage |

**Reference(s)**

Hammell KL, Dohoo IR. Mortality patterns in infectious salmon anaemia virus outbreaks in New Brunswick, Canada J Fish Dis. 2005a; 28: 639-50.

Hammell KL, Dohoo IR. Risk factors associated with mortalities attributed to infectious salmon anaemia virus in New Brunswick, Canada J Fish Dis. 2005b; 28: 651-61.

### Brief description

Following the introduction of infectious salmon anemia virus to the Bay of Fundy (Canada), an epidemiological investigation of risk factors for the disease was initiated. At the time the study was started, the etiology of the mortalities was not known and cages were designated as 'outbreaks' or not, based on the pattern of mortalities observed in the cage. A large number of risk factors were evaluated and this dataset consists of the records for 182 cages which had complete data on a subset of those factors (see list below). While the factors listed below were all fixed factors (*ie* didn't change during the study period), the data were used to compute a time-varying factor: whether or not there had been another positive cage (net-pen) at the site. This was used in survival models of the time to the occurrence of an outbreak.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| sitepen | (1000*site)+cage identifier | |
| site | site identifier | |
| net-pen | cage identifier | |
| datestrt | date fish first put in cage | |
| apr01_97 | April 1 1997 | |
| date | date of outbreak OR censoring | |
| case | case (outbreak) | 0/1 |
| cummrt96 | cum. mort. during 1996 | |
| size | cage size | 0 = <10,000<br>1 = >10,000 |
| par | initial population at risk in cage (number of fish) | |
| numcage | number of cages at site | |

## isa_test

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Carol McClure, Larry Hammell | cross-sectional | 1,071 | fish |

### Reference(s)

McClure CA, Hammell KL, Stryhn H, Dohoo IR, Hawkins LJ. Application of surveillance data in evaluation of diagnostic tests for infectious salmon anemia Dis Aquat Organ. 2005; 63: 119-27.

### Brief description

Following the identification of the infectious salmon anemia virus in the Bay of Fundy (Canada), a lot of fish were tested using a variety of diagnostic tests. It was realised that tests often gave conflicting results and the available data were used to provide a preliminary evaluation of the operating characteristics of each test. Fish that were derived from sea cages (net-pens) that had a confirmed outbreak of ISA were considered to be 'gold standard positive.' Fish sampled from sites which did not have any outbreaks of ISA (in any cages) during the study period were considered 'gold standard negative.' Other fish sampled were not included in this study. Test results from a total of 1071 fish that had multiple tests performed and which could be classified as positive (*n*=264) or negative (*n*=807) were included in the dataset.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| id | case identification | |
| date | submission date | |
| site | site identification | |
| cage | cage identification | |
| subm | submission identification | |
| fish | fish number for each case | |
| dz | disease status (clinical) | |
| histo | histology | 0 = negative<br>1 = suspicious<br>2 = positive |
| histo_np | histo neg/pos (pos=susp+pos) | 0/1 |
| ifat1 | IFAT laboratory 1 | 0-4 |
| ifat1_np | IFAT-lab1 neg/pos (pos is 1) | 0/1 |
| ifat2 | IFAT - laboratory 2 | 0-4 |
| ifat2_np | IFAT-lab2 neg/pos (pos is 2) | 0/1 |
| pcr | PCR | 0/1 |
| vi | virus isolation | 0/1 |

## ketosis

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| na | single cohort (hypothetical) | 617 | cow |

**Reference(s)**

na

**Brief description**

This is a hypothetical dataset for evaluating the condition of a cow at calving on the risk of clinical ketosis. Cows' body condition were evaluated at calving and cows were followed for 3 months to observe for clinical ketosis.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| bcs | body condition score | 0 = normal (bcs<4)<br>1 = fat (bcs>=4) |
| ketosis | clinical ketosis | 0/1 |

### lympho and lympho_mo

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Ian Dohoo | clinical trial (hypothetical) | 300 | dog |

**Reference(s)**

na

**Brief description**

These data are from a fictional clinical trial of two treatments for lymphosarcoma in dogs. The study was (hypothetically) conducted as a multicentre (*n*=10 clinics) controlled trial. Dogs meeting the eligibility criteria for entry into the trial (*n*=2000) had the tumour surgically removed (only dogs with tumours which could be surgically removed were eligible) and then were randomly assigned to one of four treatment groups: no treatment, radiation only, chemotherapy only and both radiation and chemotherapy. Dogs were randomly assigned within each centre, so the total number of dogs on each treatment group are not exactly equal for all treatments. Each dog was followed from the time of treatment until it died from a relapse of the lymphosarcoma or was lost to follow-up (*eg* died of other causes, owner moved away from the study site) and the time to the occurrence of either of those was recorded.

**Table of variables—lympho**

| Variable | Description | Codes/units |
|---|---|---|
| id | dog identification number | |
| clinic | clinic identification number | |
| age_dx | age at diagnosis (years) | yrs |
| rad | radiation therapy | 0/1 |
| chemo | chemotherapy | 0/1 |
| died | died from lympho or censored | 0 = censored<br>1 = died |
| days | time from dx to death (or censoring) | days |

The same variables are present in lympho_mo but in this data set the time variable is -months- instead of -days-.

### meta_parasite

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Javier Sanchez | meta-analysis | 75 | groups of cows |

**Reference(s)**

Sanchez J, Dohoo I, Carrier J, DesCôteaux L. A meta-analysis of the milk-production response after anthelmintic treatment in naturally infected adult dairy cows Prev Vet Med. 2004; 63: 237-56.

**Brief description**

A systematic review of the literature dealing with the effect of anthelmintic treatment on milk production in dairy cattle was carried out. Data on 75 trials were retrieved and these trials used a wide variety of measures of milk production. There was also considerable variablitiy in how the treatments were administered and other aspects of trial design.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| std_num | study number | |
| pubyear | publication year: | |
| study_lbl1 - 3 | 3 different ways of labelling the study for various graphs | |
| mlkmeas | milk production measure | 12 categories |
| tx_n | sample size treated group | |
| tx_mean | treatment response | |
| tx_sd | S.D. treatment response | |
| ctrl_n | sample size placebo group | |
| ctrl_mean | placebo response | |
| ctrl_sd | S.D. placebo response | |
| md_n | sample size | |
| md | treatment effect | |
| md_se | SE (mean diff) | |
| md_lci | lower CI (mean diff) | |
| md_uci | upper CI (mean diff) | |
| rand | the clinical trial was randomized | 0/1 |
| trblind | treatment administration blinded | 0/1 |
| clrout | clearly defined outcome | 0/1 |
| critpar | eligibility criteria for participants | 0/1 |
| pubtype | publication type | 1 = journal<br>2 = abstract<br>3 = paper in proceedings<br>4 = non-indexed journal |
| nbrhrd | number of farms | |
| drug | drug: | 17 categories |

| Variable | Description | Codes/units |
|---|---|---|
| hrdtx | whole herd treatment | 0/1 |
| nbrtx | number of treatments | 1-28 |
| parity | parity: | 0 = 1st<br>1 = 2nd<br>2 = all combined |
| tx_cat | treatment category | 0 = dry-calving<br>1 = mid-lactn.<br>2 = strategic |
| endecto | endectocide treatment | 0/1 |

## nocardia

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Lynn Ferns, Ian Dohoo | case-control | 108 | herd |

**Reference(s)**

Ferns L, Dohoo I, Donald A. A case-control study of *Nocardia* mastitis in Nova Scotia dairy herds Can Vet J. 1991; 32: 673-7.

### Brief description

This dataset contains a subset of the data obtained from a case-control study of Nova Scotia dairy herds with and without *Nocardia* mastitis. There had been a dramatic increase in the incidence of *Nocardia* mastitis in Canada since 1987 and this study was carried out to identify risk factors associated with the occurrence this disease. A total of 54 case herds and 54 control herds were visited for data-collection purposes during the summer of 1989.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| id | herd identification number | |
| casecont | case/control status of herd | 0 = control<br>1 = case |
| numcow | number of cows milked | |
| prod | average milk production for the herd | kg/cow/day |
| bscc | average bulk-tank SCC over the first 6 months of 1988 | '000s of cells/ml |
| dbarn | type of barn dry cows kept in | 1 = freestall<br>2 = tiestall<br>3 = other |
| dout | type of outdoor area used for dry cows | 1 = pasture<br>2 = yard/drylot<br>3 = none<br>4 = other |
| dcprep | method of teat end preparation prior to dry cow therapy administration | 1 = no prep.<br>2 = washed only<br>3 = washed and disinfected<br>4 = dry cow therapy not used |
| dcpct | percent of dry cows treated with dry-cow therapy | % |
| dneo | dry-cow product containing neomycin used on farm in last year | 0/1 |
| dclox | dry cow product containing cloxacillin used on farm in last year | 0/1 |
| doth | Other dry cow products used (eg penicillin or novobiocin based) used on farm in last year | 0/1 |

## pgtrial

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Jeff Wichtel | clinical trial | 319 | cow |

### Reference(s)

none

### Brief description

A clinical trial of the effect of prostaglandin administration at the start of the breeding period was carried out in three North Carolina dairy herds. On each of the three farms, the producer determined when he was ready to start breeding cows in his herd and at that time, cows were randomly assigned to receive a single injection of prostaglandin or a placebo. These cows were then followed (up to a maximum of 346 days) until they conceived (confirmed by rectal examination) or were culled. In addition to evaluating the effect of treatment on reproductive performance, three other factors were considered (parity, body condition score and herd).

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| herd | herd identification number | |
| cow | Cow identification number | |
| tx | treatment | 0/1 |
| lact | lactation number | |
| thin | body condition | 0 = normal<br>1 = thin |
| dar | days at risk | days |
| preg | pregnant or censored | 0 = censored<br>1 = pregnant |

### pig_adg

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Theresa Bernardo | cross-sectional | 341 | pig |

**Reference(s)**

Bernardo TM, Dohoo IR, Donald A. Effect of ascariasis and respiratory diseases on growth rates in swine Can J Vet Res. 1990; 54: 278-84.

**Brief description**

These are data on the growth performance and abattoir findings of pigs from a selection of Prince Edward Island, Canada farms. The data were collected to study the inter-relationships among respiratory diseases (atrophic rhinitis and enzootic pneumonia), ascarid levels and daily weight gain. Atrophic rhinitis score was determined by splitting the snout and measuring the space ventral to the turbinates. An adjustment to the score was made if the nasal septum was deviated. Lung scores were recorded on a scale of 0 to 3 (negative to severe pneumonia) and then converted to either the presence or absence of pneumonia. Parasite burdens were evaluated using fecal egg counts, counts of adult worms in the intestine and visual assessment of the liver for ascarid tracks. Production data were recorded by monitoring the pigs on the farms of origin from birth through to slaughter.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| farm | farm identification number | |
| pig | pig identification number | |
| sex | sex of the pig | 0 = female<br>1 = castrate |
| dtm | days to market (ie from birth to slaughter) | days |
| adg | average daily weight gain | gm |
| mm | measurement of snout space | mm |
| ar | atrophic rhinitis score | 0-5 |
| lu | lung score for enzootic pneumonia | 0 = negative<br>1 = mild<br>2 = moderate<br>3 = severe |
| pn | pneumonia (lu>0) | 0/1 |
| epg5 | fecal gastrointestinal nematode egg count at time of slaughter | eggs/5 gm |
| worms | count of nematodes in small intestine at time of slaughter | |
| li | liver score (based on number of parasite induced ‘white spots’) | 0 = negative<br>1 = mild<br>2 = severe |
| ar2 | severe atrophic rhinitis (ar>4) | 0/1 |

## pig_farm

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Dan Hurnik | cross-sectional | 69 | farm |

**Reference(s)**

Hurnik D, Dohoo I, Bate L. Types of farm management as risk factors for swine respiartory disease Prev Vet Med. 1994a; 20: 147-57.

Hurnik D, Dohoo I, Donald A, Robinson N. Factor analysis of swine farm management practices on Prince Edward Island Prev Vet Med. 1994b; 20: 135-46.

**Brief description**

A cross-sectional study of pig farms in Prince Edward Island (Canada) was carried out to investigate risk factors for respiratory diseases (enzootic pneumonia and pleuritis). The prevalence of each disease was determined at slaughter from routine evaluations of thoracic viscera. Data on risk factors were collected by the investigator during visits to each farm. Data on a wide variety of factors were collected and the challenge was to sort out relationships among these factors and between them and the respiratory diseases given a very limited sample size.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| farm_id | farm identification | |
| pneu | pneumonia prevalence | |
| pncode | pneumonia - categorical (3 levels) | 0 < 10%<br>1 = 10-40%<br>2 > 40% |
| pleur | pleuritis prevalence | |
| plcode | pleuritis - categorical (3 levels) | 0 = 0%<br>1 = 0-8%<br>2 > 8% |
| num | number of pigs examined at slaughter | |
| size | herd size | |
| growth | average daily gain | gm/day |
| cmpfd | pigs fed complete mixed feed | 0/1 |
| suppl | supplement added to feed | 0/1 |
| prmx | premix fed | 0/1 |
| strmed | starter ration medicated | 0/1 |
| selenium | selenium added to feed | 0/1 |
| dryfd | feed fed dry (vs wet) | 0 = wet<br>1 = dry |
| flrfd | pigs fed on floor | 0/1 |
| rooms | number of separate rooms in barn | |
| m3pig | air volume per pig | $m^3$ |
| shipm2 | density (pigs shipped per m2) | pigs/$m^2$ |
| exhaust | exhaust fan capacity (proportion of recommendation) | |
| inlet | air inlet size (proportion of recommendation) | |

| Variable | Description | Codes/units |
|---|---|---|
| maninlt | manual adjustment of air inlets | 0/1 |
| mixmnr | manure mixed between pens | 0/1 |
| straw | straw bedding used | 0/1 |
| washpns | frequency of pen washings (per yr) | |
| strdnst | floor space - starter hogs (sq m) | $m^2$ |
| grwdnst | floor space - grower hogs (sq m) | $m^2$ |
| fnrdnst | floor space - finishing hogs (sq m) | $m^2$ |
| lqdmnr | manure handled as a liquid | 0/1 |
| floor | floor slatted | 0/1 |
| sldprtn | solid partitions between some pens | 0/1 |
| hlfsld | half-solid partitions between some pens | 0/1 |
| pigwtr | pigs per water nipple | |
| numpen | number of pens | |
| mixgrp | pigs from multiple groups mixed | 0/1 |
| hldbck | slow growing pigs held back from slaughter | 0/1 |
| dstfrm | distance (km) to nearest hog farm | km |
| hmrsd | all pigs home raised | 0/1 |
| nmbsrc | number of sources of pigs | |
| mnlds | only minimal disease pigs raised | 0/1 |
| vet | veterinary visits per year | |
| feedsls | feed salesman visits per year | |
| neighbr | neighbour visits per year | |
| pigprdc | pig producer visits per year | |
| trucker | trucker visits per year | |
| you | owner works in barn | 0/1 |
| family | family members work in barn | 0/1 |
| hrdhlp | hired help works in barn | 0/1 |
| exprnce | years of experience | yrs |

### prew_mort

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Jette Christensen | cross-sectional | 6552 | litter |

**Reference(s)**

Christensen J, Svensmark B. Evaluation of producer-recorded causes of preweaning mortality in Danish sow herds Prev Vet Med. 1997; 32: 155-64.

**Brief description**

These data are a subset of 16 herds from a dataset collected by Jette Christensen in Denmark to study factors affecting preweaning mortality in pigs. These data have three levels in the hierarchy (litters (*n*=6552) within sows (*n*=3162) within farms (*n*=16)):

The key outcome of interest is preweaning mortality with a litter classified as having preweaning mortality or not if one or more piglets died before weaning.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| herd | unique herd id | |
| sowid | unique sow id | |
| litter | unique litter id | |
| lmort | prewmort in litter | 0/1 |
| herdtype | herd type | 0 = production<br>l = breeding herd |
| year | | |
| month | month | jan = l dec = 12 |
| quarter | quarter of year | l = jan-mar<br>2 = apr-jun<br>3 = jul-sept<br>4 = oct-dec |
| sow_parity | parity of sow | |
| sow_tx | sow required treatment (2d before to 7d after farrowing) | 0/1 |
| dead | number of dead piglets in litter | |
| lsize | litter size | |
| n | number at risk in litter | |
| stillb | number stillborn | |

## reu_cc

| **Contributor(s)** | **Study type** | **# records** | **Unit of record** |
|---|---|---|---|
| Emmanuel Tillard,<br>Ian Dohoo | single cohort | 2509 | lactation |

**Reference(s)**

Dohoo IR, Tillard E, Stryhn H, Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in Reunion Island Prev Vet Med. 2001; 50: 127-44.

**Brief description**

These data were collected as part of an ongoing research programme into dairy cattle fertility being carried out on Reunion Island (a French overseas department located in the Indian ocean) by researchers with CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement). Two separate datasets have been compiled. This one contains data about the calving to conception interval, while the second had data on the interval from calving to 1st service and the success of the 1st service. The data have a 4-level hierarchy (lactations (*n*=2509) within cows (*n*=1345) within herds (*n*=50) within geographic regions (*n*=5)).

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| region | geographic region | |
| herd | herd number | |
| cow | unique cow number | |
| obs | unique observation number | |
| lact | lactation number | |
| cc | calving to conception interval | days |
| lncc | calving to conception interval - log transformed | |
| lncfs_ct | calving to first service interval - log transformed and centred | |
| heifer | age | 0 = multiparous<br>1 = primiparous |
| ai | type of insemination at first service | 0 = natural<br>1 = ai |

## reu_cfs

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Emmanuel Tillard,<br>Ian Dohoo | single cohort | 3027 | lactation |

**Reference(s)**

Dohoo IR, Tillard E, Stryhn H, Faye B. The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle in Reunion Island Prev Vet Med. 2001; 50: 127-44.

**Brief description**

Same study as reu_cc except this dataset contains data on the interval from calving to first service and whether or not the first service resulted in conception. The data have a 4-level hierarchy lactations (*n*=3027) within cows (*n*=1575) within herds (*n*=50) within geographic regions (*n*=5)].

A second dataset containing only the first recorded lactation within each cow was saved as reu_cfs_1lact.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| region | geographic region | |
| herd | herd number | |
| cow | unique cow number | |
| obs | unique observation number | |
| lact | lactation number | |
| cfs | calving to first service interval | days |
| lncfs | calving to first service interval - log transformed | |
| fscr | first service conception | 0/1 |
| heifer | age | 0 = multiparous<br>1 = primiparous |
| ai | type of insemination at first service | 0 = natural<br>1 = ai |

### sal_outbrk

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Tine Hald | matched case-control | 112 | individual (person) |

**Reference(s)**

Molbak K, Hald D. An outbreak of *Salmonella typhimurium* in the county of Funen during late summer. A case-controlled study Ugeskr Laeger. 1997; 159: 36.

**Brief description**

The data are from an investigation of an outbreak of Salmonella in Funen County of Denmark in 1996. The data consisted of 39 cases of Salmonella typhimurium phage type 12 and 73 controls matched for age, sex and municipality of residence. Data on numerous food exposures were recorded and a small subset of those data are included in the dataset -sal_outbrk-.

**Table of variables**

| Variable | Description | Codes/units |
|---|---|---|
| match-grp | case-control pair identifier | |
| date | interview date | |
| age | age | yrs |
| gender | gender | 0 = male<br>1 = female |
| casecontrol | case-control status | 0/1 |
| eatbeef | ate beef in previous 72 hours | 0/1 |
| eatpork | ate pork in previous 72 hours | 0/1 |
| eatveal | ate veal in previous 72 hours | 0/1 |
| eatlamb | ate lamb in previous 72 hours | 0/1 |
| eatpoul | ate poultry in previous 72 hours | 0/1 |
| eatcold | ate cold sliced meats in previous 72 hours | 0/1 |
| eatveg | ate vegetables in previous 72 hours | 0/1 |
| eatfruit | ate fruit in previous 72 hours | 0/1 |
| eateggs | ate eggs in previous 72 hours | 0/1 |
| slt_a | ate pork processed at slaughterhouse A | 0/1 |
| dlr_a | ate pork marketed by wholesaler A | 0/1 |
| dlr_b | ate pork marketed by wholesaler B | 0/1 |

## scc_40

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Jens Agger and Danish Cattle Organization, Paul Bartlett, Henrik Stryhn | longitudinal | 14,357 | test-day observations |

### Reference(s)

Stryhn H, Andersen JS, Bartlett PC, Agger JFA. Milk production in cows studies by linear mixed models. Proc. of symposium in applied statistics, Copenhagen, January 2001. Proceedings (ed. Jensen NE. Linde P): 1-10.

### Brief description

These data are a very small subset of a large mastitis dataset collected by Jens Agger and the Danish Cattle Organization. This dataset contains records from 14,357 test-day observations in 2,178 cows from 40 herds. Milk weights (production records) were collected approximately monthly, and only records from a single lactation for each cow were included in this dataset. Factors that may have affected the somatic cell count (SCC) were also recorded. The major objective of this study was to determine if the relationship between the somatic cell count and milk production varies for cows with different characteristics (age, breed, grazing or not *etc*).

A subset of these data called -scc40_2level- was created by only taking the first observation for each cow, thereby reducing the dataset to two levels (herds and cows).

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| herdid | herd id | |
| cowid | cow id | |
| test | approximate month of lactation | 0 to 10 |
| h_size | average herdsize | |
| c_heifer | parity of the cow | 1 = heifer<br>0 = multiparous |
| t_season | season of test day | l = jan-mar<br>2 = apr-jun<br>3 = jul-sep<br>4 = oct-dec |
| t_dim | days in milk on test-day | days |
| t_lnscc | log somatic cell count on test day | |

## scc_heifer

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Sarne de Vliegher | single cohort | 10996 | test-day observation |

### Reference(s)

De Vliegher S, Barkema HW, Stryhn H, Opsomer G, de Kruif A. Impact of early lactation somatic cell count in heifers on somatic cell counts over the first lactation J Dairy Sci. 2004; 87: 3672-82.

### Brief description

These data are a small subset of a dataset on heifer collected in Belgian herds during the years 2000-2001. The objective of the research was to explore and quantify the predictive power of measuring an early lactation (days in milk 5-14) somatic cell count for milk yield and quality throughout the lactation. The heifers were followed by approximately monthly recordings until dry-off. This dataset however includes only a single record for each heifer obtained during 76-105 days in milk, for a total of 10,996 heifers in 3095 herds. The predictor of primary interest was thus the (natural) log SCC in early lactation (-lnsccel-), and the outcome of interest was the measured log SCC (-lnscc-) within the time frame 76-105 days.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| herd | herd id | |
| cow | cow id | |
| lnsccel | early lactation log somatic cell count (centred) | |
| lnscc | test day log somatic cell count (from the period 76 – 105 days) | |

### stdize, stdize_dir, stdize_ind

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Ian Dohoo | cross-scetional<br>(hypothetical) | 4 | frequency counts |

**Reference(s)**
na

**Brief description**

These three small datasets are hypothetical data used to demonstrate direct and indirect standardisation of rates in Chapter 4.

**Table of variables (stdize)**

| Variable | Description | Codes/units |
|---|---|---|
| region | region | |
| type | cattle type | |
| cases | # of herds found with tb | |
| hy | herd-years at risk | |

### tac_mca

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Ahmed ElMoselmany,<br>Greg Keefe | single cohort | 69 | herd |

**Reference(s)**

Elmoslemany AM, Keefe GP, Dohoo IR, Jayarao BM. Risk factors for bacteriological quality of bulk tank milk in Prince Edward Island dairy herds. Part 1: overall risk factors J Dairy Sci. 2009a; 92: 2634-43.

Elmoslemany AM, Keefe GP, Dohoo IR, Jayarao BM. Risk factors for bacteriological quality of bulk tank milk in Prince Edward Island dairy herds. Part 2: bacteria count-specific risk factors J Dairy Sci. 2009b; 92: 2644-52.

**Brief description**

Bulk tank raw milk quality was evaluated on all Prince Edward Island dairy herds (n = 235) over a 2-yr period (March 2005 to March 2007). Biweekly total aerobic bacterial counts (TAC) were conducted using a Petrifilm culture system. A herd was classified as a TAC case when the herd had at least 4 high TAC, counts out of the last 6 analyses for each test, respectively. Control herds had low counts in the last 6 analyses for each test in the corresponding case group. Case and control herds were visited by trained technicians and a wide range of herd level data relevant to udder health and milk quality was recorded. Only a select subset of variables (those significant in the original analysis) have been retained in this data set.

**Table of variables (stdize)**

| Variable | Description | Codes/units |
|---|---|---|
| tac | case or control based on total aerobic count | 0 = control<br>1 = case |
| X1 | predipping of teats (risk factor = not predipping) | 0/1 |
| X2 | clipping udder hair (risk factor = not clipping) | 0/1 |
| X3 | washing vs dry-wiping of udders (risk factor = washing) | 0/1 |
| X4 | teat end cleanliness (risk factor = dirty) | 0/1 |
| X5 | cow (udder flank and leg) hygiene (risk factor = dirty) | 0/1 |
| X6 | alkalinity of pipeline wash water (risk factor = high alkalinity) | 0/1 |

## tb_real

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Ian Dohoo, Fonda Munroe | retrospective cohort | 134 | animal groups |

**Reference(s)**

Munroe FA, Dohoo IR, McNab WB. Estimates of within-herd incidence rates of Mycobacterium bovis in Canadian cattle and cervids between 1985 and 1994 Prev Vet Med. 2000; 45: 247-56.

Munroe FA, Dohoo IR, McNab WB, Spangler L. Risk factors for the between-herd spread of Mycobacterium bovis in Canadian cattle and cervids between 1985 and 1994 Prev Vet Med. 1999; 41: 119-33.

### Brief description

A retrospective evaluation of all (*n*=9) outbreaks of tuberculosis in domestic animals (dairy and beef cattle, cervids and bison) in Canada between the years of 1985 and 1994 was carried out to investigate risk factors for the spread of tuberculosis within and between herds. Detailed records from the epidemiologic investigation of all outbreaks (including records on all contact herds) were reviewed and a summary of each outbreak prepared. This dataset contains data only from herds in which tuberculosis was observed. In each herd, the most probably date on which the infection entered the herd was determined and the number of new cases arising within the herd determined from the herd testing results. The number of animals in each age, sex and type group was determined and the number of animal days at risk was computed. The effects of age (three groups), sex (two groups), and animal type (five groups) on the incidence rate of new infections was investigated. **Note** To meet confidentiality and regulatory concerns, these data have deliberately been falsified.

### Table of variables

| Variable | Description | Codes/units |
|---|---|---|
| obs | observation number | |
| farm_id | farm identification | |
| type | type of animal | 1 = dairy cattle<br>2 = beef cattle<br>3 = cervid<br>4 = other |
| sex | sex | 1 = female<br>2 = male |
| age | age category | 1 = 0-12 mo<br>2 = 12-24 mo<br>3 = >24 mo |
| reactors | number of pos/reactors in the group | |
| par | animal days at risk in the group | |

### vietnam (various files)

| Contributor(s) | Study type | # records | Unit of record |
|---|---|---|---|
| Dirk Pfeiffer | surveillance data | 134 | animal groups |

**Reference(s)**

Pfeiffer DU, Minh PQ, Martin V, Epprecht M, Otte MJ. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data Vet J. 2007; 174: 302-9.

**Brief description**

Data were collected through regular surveillance for outbreaks of highly pathogenic avian influenza serotype H5N1 amongst domestic poultry in a region of northern Vietnam at the administrative level of commune between 2004 and 2006. A variety of files (some in Stata and some in R format) are provided in a zipped folder. An example of the contents of one file is shown below.

**Table of variables - viet_commune_centroid**

| Variable | Description | Codes/units |
|---|---|---|
| id | commune ID | |
| x_coord | X coordinate of commune centroid point location | |
| y_coord | Y coordinate of commune centroid point location | |
| infected | commune infected with AI | |
| inf_2003_4 | commune infected with AI in 2003/04 | |
| inf_2004_5 | commune infected with AI in 2004/05 | |
| inf_2005_6 | commune infected with AI in 2005/06 | |

Some of the other data files include:

- viet_district_centroid - This geographic data layer represents commune level data aggregated at district level.
- viet_district_poly - This geographic data layer represents commune level data aggregated at district level.
- viet_region_poly - This geographic data layer represents the boundary of the region in northern Vietnam for which the data was analysed.