CHAPTER 2

METHODOLOGY AND MATERIALS

1. PROJECT DESIGN

1.1 Study site (from IDEAL Protocols Version 17, 2008)

The study site selected is an area of Western Kenya which was considered representative of smallholder livestock farming in East Africa. The focus of the study was on the sedentary mixed crop-livestock smallholding system which encompasses >50% of poor people resident in East Africa. The area was approximately 45 x 90km and covered some or all of Busia (95.9%), Teso (96.3%), Siaya (55.5%), Butere/Mumias (26.9%) and Bungoma (20.4%) sub-locations. A sub-location is the smallest administrative unit and is typically about 10 km across and contains ∼ 80–90 households / km^2 with plots 1-5 ha in size, and with ∼ 60% households owning cattle. A typical household might own 3 cattle – one cow, one follower and one calf – although the range may be from 1 to 10, and the animals would normally be grazed communally. The main land use in the area is cultivation of maize, sugarcane, cotton, pigeon-peas and sisal. Cattle are used for animal traction, milk and meat, and the dung for fertilizer. In a survey on animal husbandry done in Busia, the main grazing system (60.4% of farmers) in the district was restricted grazing by tethering and the majority of farmers (79.9%) kept their cattle in an animal house at night (Jaetzold & Schmidt 1983).

1.2 Survey design (from IDEAL Protocols version 17, 2008)

The survey design can be described as a "stratified 2-stage cluster sample". Stratification was by AEZ. Cattle density data were believed to be aggregated at a stratum one level up from the sub-location stratum. Agro-ecological zones that were represented in the study area were Lower midlands 1 (LM1), LM2 (with a northern, middle and southern section), LM3, LM4 and Upper midlands 3 (UM3) (Figure 2.1). It was anticipated that the random cluster sampling would ensure that there was reasonable representation of farmers with varying herd sizes and management systems. The 1st stage cluster (sub-location selection) was selected using stratified-random sampling with replacement. The 2nd stage (calf selection) used ordinary random sampling without replacement, with a total sample size of 548.

Figure 2 Proportional random sample of sub-locations within AEZs [Total = 20; Random sample with replacement].
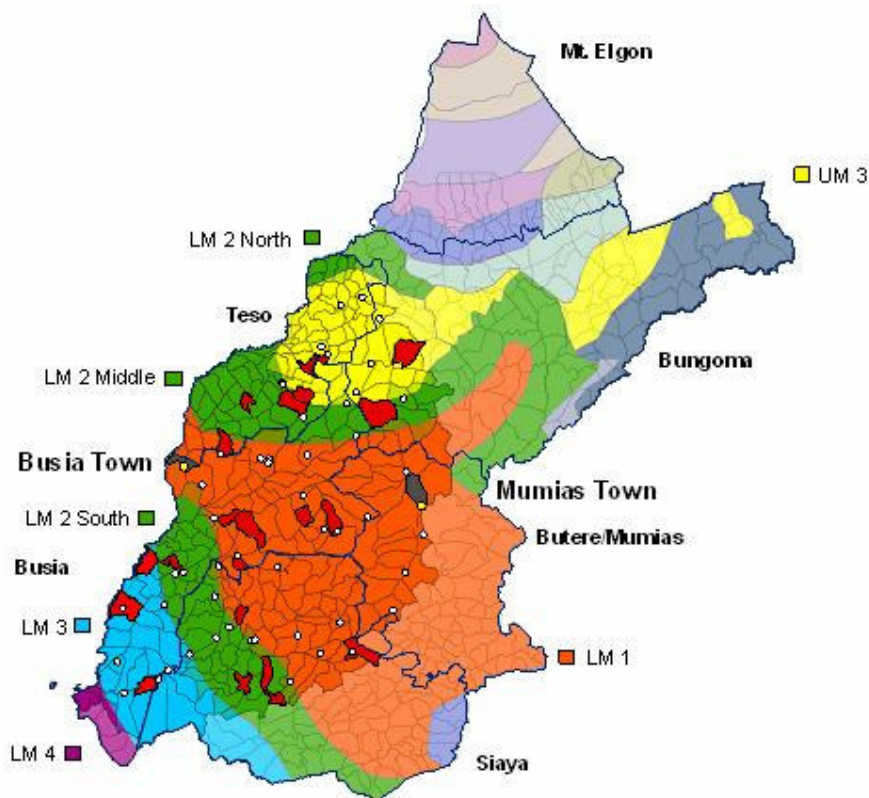
Trading/market centres are represented by white dots [n = 50]. Townships are represented by yellow dots [n = 2]. Excluded sub-locations [those containing Busia and Mumias townships] are shown in grey. Randomly selected sub-locations are shown in red.

Selection of sub-locations (from IDEAL Protocols version 17, 2008)

Sub-locations within each AEZ were selected using random sampling, with replacement. The proportion of sub-locations sampled in each AEZ was equal to the proportion of sub-locations of each AEZ in the total survey area. The minimum sample size within a sub-location was set at 3 (see Table 2.1).

The criteria to include a sub-location within the data frame for subsequent random selection (see Table 2.2) was that its geographical centre lay within 45 km of Busia town [n = 287 out of 306 sub-locations]. Mjini and Matungu sub-locations in LM1 were in Busia and Mumias townships respectively and were excluded. LM4 and LM2 north were not included as both would have resulted in a sub-location sample selection of <3.

Table 2.1 Weighted selections of sub-locations within AEZ

AEZ	Number of sub-locations in the AEZ*	Proportion of sub-locations in AEZ	Number of sub-locations to sample
LM1	[116 – 2 excluded sub-locations] = 114	0.40	8
LM2	86 [5 north; 37 middle; 44 south]	0.30 [0.02 north; 0.13 middle; 0.15 south]	6 [north not included; 3 middle; 3 south]
LM3	28	0.10	3
LM4	4	0.01	Not included
UM3	53	0.19	3
Totals	287– 2 = 285		20

* AEZ Sub-locations within 45 km of Busia

Selection of calves (from IDEAL Protocols version 17, 2008)

After receiving reports of newly born calves from farmers and local animal-health technicians, farmers/calves were randomly selected. All farmers in a sub-location were initially eligible for recruitment, on condition that the household was prepared to allow close monitoring of the calf. Once a calf was recruited from a farmer, the farmer was not eligible for selection in subsequent recruitment periods. That implied sampling without replacement. However, the population was large enough that the exclusion of previous recruited farmers/calves had little effect on selection probability or sampling variability.

Exclusion criteria for calves (from IDEAL Protocols version 17, 2008)

The traditional smallholder livestock-keeping farmers of South Western Kenya keep indigenous Zebu cattle. Calves needed to be representative of the area, and needed to have been exposed to the infectious pathogens normally occurring in the area.

A calf was not recruited in the study if:

- The dam had spent <12 months within the interest herd, regardless of its previous origin.
- The calf was born from artificial insemination.
- The calf was born with a congenital disorder.
- The farmer practised stall feeding only. It was expected that exclusion of such farmers would result in the exclusion of the vast majority of 'purely exotic' animals in the study.

Table 2.2. List and details of 20 randomly selected sub-locations

Data Source: last 'Kenya Official Map of Political Boundaries [1999]' & last 'Kenya Official Human Population Census [1999]'.

AEZ	Sub-location	Location	Division	District	Number of households	Area [km^2]
UM3	East Siboti	Siboti	Bumula	Bungoma	1245	15.8
	Kokare	Kokare	Amagoro	Teso	325	8.29
	Kidera	Kotur	Amukura	Teso	314	7.36
LM1	Yiro West	South Ugenya	Ugunja	Siaya	1361	13.7
	Simur East	Ukwala	Ukwala	Siaya	415	4.32
	Igero	Lwanya	Matayos	Busia	532	5.60
	Bumala 'A'	Bumala	Butula	Busia	724	4.38
	Ikonzo	Bujumba	Butula	Busia	1421	16.4
	Bulwani	Elugulu	Butula	Busia	478	6.87
	Bukati	Elukhari	Butula	Busia	993	11.2
	Otimong	Ochude	Chakol	Teso	506	8.66
LM2 MIDDLE	South Myanga	Kimatuni	Bumula	Bungoma	1575	22.5
	Kamunuoit	Kaujakito	Amukura	Teso	556	11.0
	Karisa	Apegei	Chakol	Teso	292	4.63
LM2 SOUTH	Kondieri	West Alego	Uranga	Siaya	630	6.38
	Namboboto	Namboboto Central	Funyula	Busia	351	4.46
	Ojwando 'B'	Alego	Boro	Siaya	832	12.6
LM3	Luanda	Namboboto	Funyula	Busia	726	9.76
	Bujwanga	Nanguba	Funyula Central	Busia	1025	16.7
	Magombe	East Bunyala	Budalangi	Busia	578	7.67

2. MONITORING AND SAMPLING OF CALVES

2.1 Recruitment of calves

Calves were recruited to the study starting October 2007 and ending September 2009. Farmers reported births of calves to either village elders or sub-chiefs in each sub-location who then reported back to the IDEAL Office, Busia. IDEAL technicians then randomly selected the required number of calves from the total number of calves reported in that week for each sub-location. The sample from which the random selection was made only included calves reported 3-7 days before the due day of the visit. Therefore, the calves had to be no less than 3 days and no more than 7 days old when first visited.

The farmer was interviewed and the protocol of the IDEAL project was explained. A signed contract agreement was reached and the farmer was paid a market-related price for the calf.

At recruitment a thorough herd history was taken and a clinical examination was done on the calf, including weight. The calf and dam were both ear-tagged with a unique barcode that had reference to the calf identification, sub-location and AEZ. Data were recorded on a visit record form as well as electronically in a personal digital assistant (PDA). Routine samples were taken (described under Sampling). Samples of the dam and photographs of both dam and calf were taken for morphometrics (not part of this study).

2.2 Monitoring

Calves were routinely visited by the IDEAL technicians, starting from week 1 (recruitment), every 5 weeks, up to 51 weeks of age. This implies that a calf that survived to 51 weeks of age, was routinely visited 11 times in total over its follow-up period. At each visit a clinical examination of the calf was performed, including a FAMACHA© score, and samples (see 2.3) were collected. Clinical episodes outside the frame of 5-weekly monitoring were reported by either the farmer or the local government Animal Health Technician. The IDEAL team then visited the calf the following day. Monitoring of calves started 12 November 2007 and ended 30 August 2010 when the last recruited calves were 51 weeks old.

2.3 Sampling

(Test methodology is discussed under Diagnostics methods)

Routine sampling occurred at recruitment of the calf and then every five weeks up to 51 weeks of age. That implied sampling each calf 11 times. Samples taken at routine visits, including recruitment visits, were the following:

- Two thin peripheral blood smears (MNS) collected from the marginal ear vein for manual haematology and screening of blood-borne protozoa. Blood was collected with a micro-capillary tube and decanted onto a glass slide.
- Two thick peripheral blood smears (MKS) collected from the marginal ear vein for screening of blood-borne protozoa. Blood was collected with a micro-capillary tube and decanted onto a glass slide.
- Jugular blood in one 10 ml plain vacutainer tube (RPL) for the separation of serum. At the 51-week visit two RPL samples were taken.
- Jugular blood in two 5ml EDTA plastic tubes (RED) for haematology, screening of trypanosomes and molecular diagnostics.
- Faecal samples (FCS) from the rectum for screening for helminths and protozoa.
- Tick species observed to be present on the calf were reported on the visit form.

Calves were visited outside the frame of routine visits whenever a clinical episode was reported. In the case of clinical episodes, whether it coincided with routine visits or not, the same samples were taken as above, as well as the following:
- Jugular blood in one 10 ml heparin tubes (SHB) for molecular diagnostics, virology or toxicology (not part of this study).
- Additional samples as needed, such as fine-needle aspirates of enlarged lymph nodes.

3. DIAGNOSTIC PROCEDURES

3.1 Sample processing and testing
Initial sample processing occurred at the local ILRI laboratory, Busia. Where samples from calves were taken in duplicate, one sample was processed further as described hereafter, and one sample was stored in the biobank at ILRI, Nairobi.
- One set of MNS and MKS was stained with Giemsa 10% for 30 minutes for microscopy for blood-borne protozoa.
- RPL samples were spun down and decanted. Two aliquots were made from serum (four aliquots for 51 week visits). One aliquot of 100 µL was used to measure total serum protein (TSP) with a refractometer (model RHC-200ATC, Westover Scientific). Aliquots were then stored at -20˚C prior to shipment to ILRI for serology.
- One of each RED set was stored at -80˚C prior to shipment to ILRI for PCR and RLBT. Aliquots were stored in the Biobank. One of each RED was processed at

Busia as follows: Haematology was done using the automated blood cell analyser, pocH-100iV Diff (Sysmex© Europe GMBH); two thin (BNS) and two thick blood smears (BKS) were made to check for blood-borne parasites; and two heparinised microcapillary tubes were filled and centrifuged. The microhaematocrit was read using the microcapillary tubes with a Hawksley haematocrit reader and the sample was screened for trypanosomes (HCT) after which the buffy coat was decanted onto a glass slide for dark ground examination (DG).

- Faecal samples were tested with the McMasters test for egg counts and oocyst counts, direct Baermans test for *Dictyocaulus* spp., and sedimentation for trematodes. If the McMasters test was negative, faecal flotation was also done to confirm a negative result. If either the McMasters test or the flotation technique was positive, larval culture for nematodes was then done.

After shipment to ILRI, Nairobi, the samples were either put in storage in the biobank or subjected to further diagnostic testing, including serology for specific tick-borne diseases and polymerase chain reaction (PCR) for trypanosomes,

Duplicates of BNS were shipped to DVTD, South Africa, for manual differential counts of blood cells for all visits.

An aliquot of the DNA extracted for the trypanosome PCR were also shipped to South Africa for use in the reverse line blot technique (RLBT) for several tick-borne diseases.

3.2 Diagnostic tests

Haematology

- Manual PCV was measured by filling a microcapillary tube with blood, centrifuging at 50 000 rpm for 5 minutes and reading the haematocrit with a Hawksley microhaematocrit reader (Jain 1993). A PCV <25% was considered anaemic.
- Differential cell counts
  Automated blood cell analysis was done with pocH-100iV Diff (Sysmex© Europe GMBH). The following parameters and units were measured by the analyser (pocH-100iV Diff User's Manual (Sysmex© Europe GMBH) 2006):
  o White blood cell count ($x10^3$/µL)
  o Red blood cell count ($x10^3$/µL)
  o Haemoglobin concentration (g/dL)

o Haematocrit (%)

o Mean corpuscular volume (MCV) (fL)

o Mean corpuscular haemoglobin (MCH) (pg)

o Mean corpuscular haemoglobin concentration (MCHC) (g/dL)

o Platelet count (x10^3/μL)

o Mean platelet volume (fL)

o Percentage: Lymphocytes (%)

o Percentage: Other white blood cell types (%)

o Absolute lymphocyte count (x10^3/μL)

o Absolute other white cells count (x10^3/μL)

In addition, manual differential counts were done on BNS samples to calculate absolute lymphocyte-, absolute eosinophil-, absolute neutrophil- and absolute monocyte counts.

Helminthology

- McMasters method (Kaufmann 1996) was used to calculate the number of nematode eggs (EPG) and number of coccidial oocysts (OPG) per gram faeces. Eggs were reported as either "strongyle" or "strongyloides" eggs. Where it was possible to identify eggs to the genus level, such as *Trichuris* sp., *Capillaria* sp. and *Nematodirus* sp., this was reported.

- Faecal flotation (Kaufmann 1996) was done if the McMasters test was negative. Results were reported as either negative or positive.

- Direct Baermann test (Kaufmann 1996) was used to identify the larvae of lungworms (*Dictyocaulus* spp.). Results were reported as either positive or negative.

- Sedimentation of faeces (Kaufmann 1996) in water was used to identify eggs of trematodes. After sedimentation the supernatant was drawn off and the sediment stained with methylene blue (5%). The eggs appear brown or yellow against a pale blue background.

- Larval culture was done as described in Kaufmann (1996). Incubation was at room temperature for 14 days after which harvesting was done using a Baermann apparatus. Identification of third-stage larvae of nematodes was done as described in Reinecke (1983). Third-stage larvae were reported as percentages of the total larvae recovered per sample.

Microscopy

Thin blood smears (MNS and BNS) were air-dried, fixed in absolute alcohol and stained for 30 minutes with Giemsa 10% dilution. Thick blood smears (MKS and BKS) were air-dried, and stained as above. Smears were examined for blood-borne parasites with a standard light microscope under 10-100x magnification.

Quantification of parasitaemia on MNS and BNS was as follows:
- 1 = one organism found in every >10 fields;
- 2 = one organism found in every >1 field and <10 fields;
- 3 = one or more organisms found per field.

Trypanosome parasite concentration techniques
- Microhaematrocrit centrifugation technique (HCT), also called the Woo method (OIE 2005) was done using the microhaematocrit tubes that were used to read the PCV. Results were reported as either negative or positive.
- Dark-ground/phase-contrast buffy coat technique (DG) (OIE 2005) followed the HCT. Where species identification was possible based on morphology and movement of the parasites, it was reported as such.

Serology

Indirect enzyme-linked immunosorbent assays (ELISA) were used to evaluate the level of antibodies to *T. parva, T. mutans, A. marginale, B. bovis,* and *B. bigemina* (Katende, Goddeeris, Morzaria, Nkonge & Musoke 1990; Katende, Morzaria, Toye, Skilton, Nene, Nkonge & Musoke 1998). The results were expressed as percent positivity (PP) values of optical density (O.D.), relative to those of a strong positive control (Wright, Nilsson, Van Rooij, Lelenta & Jeggo 1993). A PP cut-off of 20 was considered positive for *T. parva* and *T. mutans*, and a PP cut-off of 15 was considered positive for *A. marginale* and *B. bigemina*.

Molecular diagnostics
- Trypanosome PCR

  A multi-species un-nested PCR was used to test for trypanosomes *(T. brucei; T. theileri; T. congolense* savannah*; T. c.* forest*; T. c.* kilifi*; T. c.* tsavo*;* and *T. vivax)* (Thumbi, McOdimbam, Mosi & Jung'a 2008). This PCR was done on the 51-week samples, as well as samples from individual visits where a calf tested positive for trypanosomes on microscopy (HCT and/or DG).

- RLBT

  RLBT was done at 51 weeks as a screening test for *T. parva, T. mutans, T. taurotragi, T. velifera, T. buffeli, Theileria* sp.*, A. centrale, A. marginale, B. bovis, B. bigemina, A. (E.) bovis,* and *E. ruminantium* (Gubbels, De Vos, Van der Weide, Viseras, Schouls, De Vries & Jongejan 1999; Bekker, De Vos, Taoufik, Sparangano & Jongejan 2002).

4. DATA ANALYSIS

The data analysis techniques and their applications specific for that chapter are discussed in more detail in the relevant chapters. The general techniques are discussed here. The computation of the results and the production of the graphs were done using R 2.8.1 (Ihaka & Gentleman 1996). All plots, except Kaplan-Meier curves, were drawn with R package *ggplots* 0.8.9 (Wickham 2009). The *Survival* package 2.36-9 in R (Lumley 2007) was used to calculate life tables and Kaplan-Meier plots where applicable.

4.1 Descriptive statistics

Descriptive statistics used in either one or more of the results chapters included the mean and the standard error (SE). The 95% confidence intervals (95%CI) for the means were calculated as follows:

$$\text{mean} - \text{t.value*SE to mean} + \text{t.value*SE},$$

where t.value was read from the table of t distribution probability and SE was calculated as the square root of variance (var) divided by the sample size (n) (Downie & Heath 1970). Differences in means were calculated, as appropriate, with either the Student's t-test (using R function "t.test(x)"), or where assumptions of normality was not met, the Mann-Whitney Test (using R function "wilcox.test(y ~ x)"). A p-value <0.05 was considered as significant (Crawley 2007). The 95% CI was calculated as follows:

$$(\text{difference in mean}) - 1.96\text{*SE to (difference in mean)} + 1.96\text{*SE},$$

where SE is square root of the sum of $\text{var}(x_1)^2/n(x_1)$ and $\text{var}(x_2)^2/n(x_2)$.

The 95% confidence intervals for proportions (p), such as prevalences, were calculated as follows:

$$p - (1.96\text{*SE}) \text{ to } p + (1.96\text{*SE}),$$

where p is a percentage and SE is the square root of (p*(100-p)/n) (Downie & Heath 1970).

4.2 Evaluation of diagnostic tests

In Chapter 3 the test performance of the FAMACHA© score test was evaluated using the *ROCR* (1.0-4) package from R (Sing, Sander, Beerenwinkel & Lengauer 2005). The function "prediction(x)" calculated the values of a 2x2 contingency table as a first step. The function "performance(x)" was then used to calculate all of the following measures using the contingency table values: "sens": Sensitivity (Se); "spes": Specificity (Sp); "ppv": postitive predictive value (PPV+); "npv": negative predictive values (PPV-); and "auc": area under the curve (AUC). The AUC is equal to the value of the Wilcoxon-Mann-Whitney test statistic (Greiner, Pfeiffer & Smith 2000). The 95% confidence intervals for the AUC could thus be calculated by rank correlation for censored data using the Hmisc (3.8-3) package with the function "rcorr.cens" (x). The output value "C Index" was then used to calculate the confidence interval as follows:

$$\text{"C Index"} – 1.96*SE \text{ to } \text{"C Index"} + 1.96*SE.$$

4.3 Statistical modelling

Longitudinal studies allow one to investigate within-calf variation over time as well as between-calf variation. The assumption of independence of data was not met, however, since these were repeated measures of individual calves within a longitudinal study which implied clustering within calves (Finucane, Samet & Horton 2007).
.
Generalized mixed-effect (GME) models were used to investigate the association between pathogens and the haematological measure of interest. The advantage of using mixed-effect models is that it takes into account the problem of repeated measures and allows for data to be unbalanced and/or incomplete (missing values for some time points). By distinguishing between parameters that affect the population equally (fixed effects) and factors that affect the individual calf differently (random effects), the mixed-effect models can distinguish within-calf from between-calf sources of variation (Finucane *et al.* 2007).

Generalized additive mixed models (GAMM) are an extension of generalized mixed-effect models and allow the use of the non-parametric smoothers in addition to parametric components where applicable (Wood 2006). Non-parametric smoothers were used to account for the non-linear effect of age on the distribution of the haematological parameters. The smoother function is denoted in the model by "s(x)", which produces isotropic smooths of multiple predictors (Wood 2006). Modelling of GAMM was done in R using the package

mgcv (1.7-0). In mgcv, generalized cross validation (GCV) criterion or unbiased risk estimator (UBRE) are used to estimate the smoothing parameters (Crawley 2007). Models are fitted using penalized likelihood maximization (Crawley 2007). The model was fitted by the following formula, using the "gamm" function:

```
model <- gamm(fixed=response~covariates, correlation=corAR1(), random=~covariates);
```

where the "corAR1" function allowed for autocorrelation nested within calves.

Calf number (CalfID) and sublocation (SL) were included as random effects. The consistent inclusion of SL conditions the model for environment and nutrition-related factors and differences in exposure levels to pathogens (Van der Waaij *et al.* 2003). Varying slopes and intercepts models were used and temporal autocorrelation within calves was accounted for in the model specifications. The cutoff for statistical significance was set to $p < 0.05$.

Model checking was done by inspecting the fit of the model by testing for heteroscedasticity by plotting the residuals against fitted values, and testing for non-normality of errors by plotting the residuals against the standard normal deviances (Crawley 2007). Akaike's information criteria (AIC) were also used as a measure of model fit. The model with the lowest AIC was chosen as a better fit.