FACULTY OF VETERINARY MEDICINE
approved by EAEVE

Modeling the dynamics of infectious animal diseases using the frailty model

Bart Ampe

Dissertation submitted in fulfillment of the requirements
for the degree of Doctor (PhD) in Veterinary Sciences, Ghent University, 2012

Promotor: Prof. Dr. L. Duchateau

Faculty of Veterinary Medicine
Department of Comparative Physiology and Biometrics

Contents

List of abbreviations **1**

1 Overview of the relevant survival analysis techniques and the datasets **3**

1.1 Introduction 5
1.2 Infectious bovine rhinotracheitis dataset 6
1.2.1 An introduction to infectious bovine rhinotracheitis . 6
1.2.2 Different types of vaccines 8
1.2.3 Analysis of BoHV-1 vaccination studies 10
1.2.4 The dataset 11
1.3 The intramammary infection dataset 15
1.3.1 An introduction to intramammary infections and mastitis 15
1.3.2 Pathogenesis of mastitis 16
1.3.3 Mastitis causing pathogens 17
1.3.4 Diagnosis of mastitis 21
1.3.5 Analysis of intramammary infections studies 22
1.3.6 Analysis of studies about the effect of an intramammary infection with a minor pathogen 25
1.3.7 The dataset 26
1.4 Survival data 31
1.4.1 Basic notation and functions in survival analysis . . . 32
1.4.2 Censoring 33
1.4.3 The proportional hazards model 36
1.4.4 Multivariate survival data: the frailty model 38
1.5 Research objectives 41
Bibliography 42

2 Long-term effect of hyperimmunization on the transmission of infectious bovine rhinotracheitis virus in cattle herds 53
2.1 Introduction 55
2.2 Material and Methods 57
2.2.1 Animals 57
2.2.2 Procedures 57
2.2.3 Serologic monitoring 57
2.2.4 Questionnaire for the risk factor analysis 58
2.2.5 Data analysis 58
2.3 Results 60
2.3.1 Animals 60
2.3.2 Dairy herds 61
2.3.3 Dairy-beef herds 64
2.4 Discussion 67
2.5 Conclusion 69
Bibliography 70

3 Investigating clustering in interval-censored udder quarter infection times in dairy cows using a gamma frailty model 73
3.1 Introduction 75
3.2 Materials and methods 76
3.2.1 Animal database 76
3.2.2 Statistical methods 77
3.3 Results 80
3.4 Discussion 86
3.5 Conclusion 87
Bibliography 89

4 Investigating the protective effect of coagulase-negative staphylococci or *Corynebacterium bovis* infection on intramammary pathogenic infections using a frailty model based approach. 93
4.1 Introduction 95
4.2 Materials and methods 97
4.2.1 Intramammary Infections Dataset 97
4.2.2 Data Analysis 98
4.2.3 Simulation Study 100
4.3 Results 101
4.3.1 Intramammary Infections Dataset 101
4.3.2 Simulation Study 103

4.4 Discussion 104
4.5 Conclusions 106
Appendix 106
Bibliography 110

5 General discussion and future research perspectives 113
5.1 Clinical part 115
5.1.1 IBR dataset 115
5.1.2 Intramammary infections dataset 116
5.2 Statistical methodology 119
5.2.1 A piecewise constant hazard model with calendar time 119
5.2.2 Shared frailty model for interval-censored survival data with time varying covariates 121
Bibliography 124

Summary 126

Samenvatting 132

Dankwoord 138

List of abbreviations

$\boldsymbol{\beta}$	Vector of regression parameters
BoHV-1	Bovine herpesvirus 1
CNS	Coagulase-negative staphylococci
δ	Censoring indicator
$f(.)$	Density function
$F(.)$	Cumulative distribution function
gB	Glycoprotein B
gE	Glycoprotein E
γ	Shape parameter of the Weibull distribution
$h(.)$	Hazard function
$H(.)$	Cumulative hazard function
HIG	Hyperimmunization group
HR	Hazard ratio
IBR	Infectious bovine rhinotracheitis
IMI	Intramammary infection
L	Likelihood
l	Loglikelihood
λ	Scale parameter of the Weibull distribution
n	Sample size
NIG	Nonintervention group
PMN	Polymorphonuclear neutrophil
R_0	Basic reproduction number
$S(.)$	Survival function
SCC	Somatic cell count
T	Random variable representing the event time
TK	Thymidine kinase
θ	Variance of the frailties in the frailty model
τ	Kendall's τ
u_i	Frailty term
$\boldsymbol{x}$	Vector of covariates

Chapter 1

Overview of the relevant survival analysis techniques and the datasets

1.1 Introduction

In veterinary science, a lot of research in infectious diseases results in longitudinal data. To investigate infection dynamics, samples to determine the infection status are often taken at several predefined time points in order to estimate the time of infection. This results typically in time to event data which are often reduced to binary data for analysis: the infection did or did not occur during an interval or before a predefined endpoint. The big advantage of reducing to binary data is the simplicity of the analysis: logistic regression can be used to analyze the data and the resulting odds ratios are well known in veterinary epidemiology. Sometimes the data are even more reduced to the number of IMI on farm level, typically analyzed with a Poisson regression model. The first problem with the reduction of the data is the loss of information contained in the original data, the time variation. In the case that most infections occurred in the first part of the study period, the results will be the same as if most infections occurred in the last part. Survival models are more appropriate to analyze time to event data. Survival models model the evolution of the hazard of infection over time and give also an idea when most infections occurred.

The second problem with this reduction is how to handle missing values due to follow up. In a lot of field studies some animals leave the study population before the end of the study period, due to diverse reasons unrelated to the aim of the study. They can be sold, culled for some reason, they died due to another disease, etc. This is called right censoring in survival analysis. In logistic regression those animals are often excluded for further analysis while survival models still use the information those animals provide: there was no infection until they left the study population.

Another often observed problem in field studies is clustering. In most cases, the study population is housed in a number of farms. Animals housed in the same farm are more homogeneous than animals from different farms. Those animals are not fully independent because they have a lot of farm-related factors in common (e.g. housing facilities, veterinarian, farm management, infection pressure, ...). This clustering is often ignored in the statistical analysis although it is possible to solve this issue by introducing a random effect. Logistic regression can be extended by introduction of a random effect and a shared frailty term can be introduced in survival models.

The last frequently observed problem is uncontinuous monitoring. In most field studies it is not feasible to sample animals on a daily basis due to practical or financial reasons. Samples are often taken at predefined intervals

(e.g. weekly or monthly sampling). It is only known that the true infection time is contained in the interval between the last observed time without infection and the first time with the infection, known as interval censoring. In case of logistic regression, the time of infection has no effect on the results and this problem can be ignored. When survival models are used, the true time of infection is needed. This is often solved by assuming that the middle of the interval is the true infection time. In this case, the extra variability due to the unknown true infection time is ignored. Techniques that take into account interval censoring are more appropriate to analyze this type of data.

In the first two parts of this introductory chapter, an introduction to the two datasets used in this thesis is given. The third part of this chapter describes some basic concepts in survival analysis. It contains the notation and the basic models which are further extended in the context of this thesis. In the last part, the general objectives are given. The two datasets introduced in this chapter will be used in the following chapters to illustrate the developed statistical techniques. In Chapter 2 the effect of hyperimmunization on the transmission of infectious bovine rhinotracheitis (IBR) virus in cattle herds is investigated. The available techniques are extended to make use of calendar time and to correct for seasonal fluctuations. In Chapters 3 and 4 the intramammary infections dataset is analyzed. The effect of parity and the clustering on cow and farm level is estimated and the effect of an IMI with *C. bovis* and CNS on the susceptibility to a new IMI with other mastitis pathogens is modeled. For that purpose, the shared frailty model is extended to include interval-censoring and time varying covariates.

1.2 Infectious bovine rhinotracheitis dataset

1.2.1 An introduction to infectious bovine rhinotracheitis

IBR is caused by bovine herpesvirus 1 (BoHV-1) which belongs to the subfamily of the Alphaherpesvirinae, characterized by a relatively large host range, short replication cycle, and the ability to induce a latent infection of the host (Tikoo et al., 1995; Muylkens et al., 2007).

The natural entrance of BoHV-1 is the mucous membrane of either the upper respiratory tract (direct nose to nose contact or airborne transmissions on short distances) or the genital tract (direct contact at copulation or through virus contaminated semen in case of artificial insemination) (Wentinck et al., 1993). After penetration into the epithelial cells, BoHV-1 sets up a lytic replication cycle resulting in necrosis and apoptosis of the

host cells. BoHV-1 may also reduce the repair mechanism of the respiratory epithelium (Spurzem et al., 1995). The huge virus replication at the natural portal of infection results in high virus titers in the nasal mucus, responsible for the rapid spread of the infection within the herd. BoHV-1 spreads in the host by viremia which can result in abortion or even a fatal systemic infection in young seronegative calves. Viral neuroinvasion during the primary virus replication in the mucosal epithelium occurs by invasion of the nerve endings in the mucosae. In case of oro-nasal infection, usually only the first order neuron, located in the trigeminal ganglion, is infected, leading to a life-long latent infection of the peripheral nervous system (Engels and Ackermann, 1996). In case of venereal infection, the sacro-iliac ganglios is latent infected (Van Engelenburg et al., 1995). Sporadically BoHV-1 has also been isolated from the central nervous system (Engels and Ackermann, 1996).

The first immune response after mucosal infection is a non-specific inflammatory and cellular reaction which is essential for initiating the specific immune response. The specific cellular immunity is detectable from the 5th day post infection and reaches a peak after 7 to 10 days while the specific humoral immunity becomes detectable 10 days after infection. Colostral antibodies from BoHV-1 immune cows protect neonatal calves against systemic and lethal disease (Babiuk et al., 1996).

After a primary infection with BoHV-1, cattle become latent carriers. The main site of latency is the peripheral nervous system, but there is evidence that latency and reactivation take place in the germinal centers of the pharyngeal tonsils. The progeny virus phenotype and the immune status were shown to influence the reactivation. The primary immune response (after infection or vaccination) as well as the secondary immune response (after booster) are effective to inhibit virus re-excretion due to high titers of BoHV-1 neutralizing antibodies (Pastoret et al., 1979).

In most cases, BoHV-1 infections are subclinical. In calves with maternal antibodies, the BoHV-1 infection results only in discrete clinical signs. Infected seronegative calves have high fever during 4 to 5 days, sometimes accompanied by apathy and anorexia. Adult cows show a significant milk drop during the next days. Two to three days after infection, the epithelium damage at the primary replication sites can result in ocular and respiratory signs, such as red appearance of nasal mucosa, serous to mucopurulent nasal secretion, heavy breathing and cough. An infection of a seronegative cow at 4 to 8 months of gestation can result in abortion. Neonatal calves without colostral antibodies may experience multisystemic infection, mostly fatal within 4 to 5 days (Lemaire et al., 2000; Muylkens et al., 2007).

The latency reactivation cycle, after stress conditions, has a major impact on the epidemiology, and is responsible for the maintenance of BoHV-1 in the herd/population. Reactivation occurs at e.g. birth, during transport or following the introduction of heifers into a group of dairy cows (Thiry et al., 1987).

Van Malderen et al. (1987) performed a sero-epidemiological survey to investigate the seroprevalence of BoHV-1 in Belgium at farm level. Mixed samples (less than 50 animals in each sample) at farm level were analyzed and 62% of the 8285 farms scored BoHV-1 positive. More than 10 years later, during 1998, Boelaert et al. (2000) performed a similar survey in non-vaccinated farms in Flanders and concluded that 67% of the farms where BoHV-1 positive, and 36% of the individual animals. Vaccinating herds were excluded from the analysis because the use of marker vaccines was not mandatory at that time.

The control strategy for BoHV-1 is mainly based on the prevention of virus transmission by vaccination and sanitary measures (e.g. purchase of seronegative animals). Most European countries have a control strategy based on the 'DIVA' strategy (differentiating infected from vaccinated animals). This is only possible if gE deleted vaccines are used. However, the detection kit (ELISA) has a low sensitivity (around 92%) which is responsible for a high number of false negative tests. The specificity the ELISA tests is almost 100% (Kramps et al., 1996; Van Oirschot et al., 1997).

1.2.2 Different types of vaccines

BoHV-1, like all *Herpesviridae*, has a virion morphology based on a icosahedral capsid symmetry covering the DNA, surrounded by a cell-derived envelope and a tegument as protein made matrix connecting the capsid and the envelope. The envelope also includes virally encoded glycoproteins who have various functions such as attachment and entry or cell-to-cell spread. Entry into a cell requires binding of virus glycoproteins to receptors on the cell surface followed by endocytosis of the virion. Some of those glycoproteins are essential (deletion leads to a lethal mutant) and some are not essential. The BoHV-1 penetration in cells requires at least the involvement of four glycoproteins: gB, gD, gH and gL (Meyer et al., 1998). One of the non-essential proteins in BoHV-1 is glycoprotein E (gE).

In Belgium there are currently two different types of vaccines (attenuated live vaccines and inactivated vaccines) in use for BoHV-1, all of them gE deleted marker vaccines. The inactivated whole virus vaccines (which were gE positive) are not allowed anymore in Belgium. The use of gE deleted

marker vaccines makes it possible to test for antibodies against the deleted glycoproteins (in most cases gE), which will only be produced when the cow was infected with a wild field strain BoHV-1 (in case of a natural infection). The possible outcomes of the ELISA test for gE and gB antibodies are illustrated in Table 1.1. The possibility to distinguish between naturally infected or only vaccinated (only gB positive) is used in the DIVA strategy (Zhao and Xi, 2011)

Table 1.1: Interpretation of ELISA testresults for gE and gB.

gE-antibodies	gB-antibodies	interpretation
-	-	not naturally infected and not vaccinated
-	+	not naturally infected but vaccinated
+	+	naturally infected

The first group of vaccines are the attenuated live vaccines. Live gE-deleted vaccines can induce early immunity after intranasal administration, which is more efficacious than the inactivated vaccines at the early stage of a BoHV-1 outbreak (Kaashoek and Van Oirschot, 1996). The big problem with life attenuated vaccines is the risk of contamination of the vaccine with other pathogens. In 1999 there was an outbreak of bovine virus diarrhea on Dutch farms, induced by a live attenuated vaccine contaminated with bovine virus diarrhea virus type 2 (Barkema et al., 2001). On the other hand, a live vaccine, intranasally administered, can be excreted in the field with the potential emergence of highly pathogenic BoHV-1 mutant strains by recombination between the gE negative vaccine strain and field strains (Muylkens et al., 2006). There is also a combination vaccine where both gE and thymidine kinase (TK) are deleted. The TK deletion seems to reduce the latency and reactivation in the nervous system. The gE and TK deleted vaccine virus were not re-excreted in the field after dexamethasone treatment (Kaashoek et al, 1996). Recently also a gG and TK deleted virus was suggested as a promising candidate for a marker vaccine against BoHV-1 (Zhang et al., 2011)

The second group are the gE-deleted inactivated (killed virus) vaccines. They are administered intramuscularly and are safer than the attenuated life vaccines. Bosch et al. (1997) showed that the inactivated vaccines were more efficacious in reducing the virus excretion after reactivation than the

live vaccines, but they induce less protection against clinical signs and fever in cattle (Bosch et al., 1996).

There are also subunit and DNA vaccines, but those are still under development. Subunit (gD) vaccines can induce high titers of protective antibodies if an appropriate adjuvant is used. The latter is the main problem and delays their use in the field (Zhao and Xi, 2011). Toussaint et al. (2007) showed that slow-release of a DNA vaccine by diffusion, using agarose hydrogel implants, induces immunity. The technique still needs some optimizations before they can be used for livestock vaccination (reduction of the size of the implants, a more user friendly administration technique).

1.2.3 Analysis of BoHV-1 vaccination studies

A review of the methodology of 6 BoHV-1 vaccination or hyperimmunisation studies showed that the applied designs were very diverse. Most of the studies were randomized trials (Bosch et al., 1996, 1997; Kerkhofs et al., 2003; Mars et al., 2001), but two were observational (Makoschey et al., 2007; Nardelli et al., 2008). The first 3 mentioned studies were experimental studies while the last tree were field studies mainly to validate vaccination/eradication protocols. This results in different objectives and measured outcomes for those studies. The main hypothesis and the statistical analysis were also different in each study.

Bosch et al. (1996 and 1997) performed a randomized vaccination-challenge trial (3 experiments of 30 animals each) and concentrated on clinical observations (scores ranging from 0 to 3) and BoHV-1 neutralizing antibody titers of the sera. The data was analyzed using ANOVA and logistic regression. The number of the experiment was introduced in the model as a fixed effect in order to eliminate the effect of the different experiments.

Kerkhofs et al. (2003) performed an experimental study (30 cows) with 4 different vaccination protocols and a control group to assess the ability to protect cows against a challenge with BoHV-1 and reactivation after dexamethasone injection. The main outcomes were clinical signs and serological parameters (gE-seropositive, cellular immune response and virus titers). The data were analyzed using an ANOVA (for titers a log10 transformation was used). Correction for clustering was not possible because the treatments were assigned at cluster level with only one cluster for each treatment.

Mars et al. (2001) performed a larger (84 herds) double blind randomized field trial with a vaccination group (live gE-negative vaccine) and a control group (saline placebo), but a lot of herds were excluded for the final analysis (exclusion of 28 herds in control group due to a lack of events or when

the initial seroprevalence was lower than 30%). The main outcome was gE-seroconversion and based on this outcome an estimation of the basic reproduction number (R_0) was estimated. A one sided test that the basic reproduction number in the vaccination group was smaller than that of the control group was performed. The basic reproduction number of the vaccination group was significantly smaller than 1. A R_0 smaller than 1 means that an infection will die out, while R_0 larger than 1 means that the infection may become endemic.

The performance of a live marker vaccine in the field was studied by Makoschey et al. (2007) in three European countries. In this observational study, in some farms (1 in Germany, 1 in Italy, 147 in Hungary) the seroprevalence was calculated after the start of a hyperimmunisation program. The results are descriptive, a decrease of the seroprevalence was observed, but no statistical analysis of the data was reported. There were also some indications of seasonal effects on the seroprevalence (an increase during winter, decrease during summer).

1.2.4 The dataset

Design

The experimental study was conducted in 34 dairy herds and 38 dairy-beef mixed herds that were selected from a pool of 92 volunteer herds. Inclusion criteria for the herds were that farming activity was the only source of family income and that a herd comprised less than 180 cattle (due to feasibility reasons and budget restrictions). The minimal herd size in the study was 60, the median was 115 and the maximum was 172 animals, the distribution of the herd size was similar for both production types.

Within each production type, herds were randomly assigned to three groups, Hyperimmunization group (HIG) 1 (10 herds), HIG 2 (10 herds) and a non-intervention group (NIG, 16 herds) (control group) by use of a randomization procedure (a lottery procedure was followed). The production type used in the stratified randomization was based on the files received from the SANITEL-CATTLE, the central computerized database for the identification and registration of the Belgian cattle population (Ministry of Small Enterprises, Traders And Agriculture, Belgium). At the start of the study, some differences between the 'Official record' and the real type were observed and further analysis was done using the real production type.

All the cattle in HIG 1 and 2 were hyperimmunized. This consisted of an initial 2 administrations of marker vaccines (interval of 3 to 5 weeks between

administrations), which were followed by booster vaccinations at intervals of approximately 6 months. To ensure identification of the treated cattle, vaccinations were performed at the start of the study (January 1997) and then successively shortly before the start of the cattle pasture periods and as soon as possible after the start of the cattle stabling periods for a period of 28 months.

Cattle in HIG 1 initially received an intranasal administration followed by an intramuscular administration of an attenuated gE-deleted marker vaccine (Bayovac IBR-Marker vivum, at the time by Bayer AG; Rhinobovin Marker Live, Intervet, previously Hoechst). Cattle in HIG 2 initially received 2 SC administrations of an inactivated virus gE-deleted marker vaccine (Bayovac IBR-Marker inactivatum, then Bayer AG; Rhinobovin Marker inactivated, Intervet, previously Hoechst). Both hyperimmunized groups received boosters inoculations of the inactivated-virus vaccines (Bayovac IBR-Marker inactivatum, then Bayer AG; Rhinobovin Marker inactivated, Intervet, previously Hoechst), SC, at the 6-month intervals. For the NIG, farmers used their usual vaccination schedules. All cattle of appropriate age for vaccination in each herd were included in the study. Only gE-deleted marker vaccines were used for all 3 groups, which allowed for serologic differentiation between infected and uninfected but vaccinated cattle.

Serologic monitoring

To limit the amount of animal handling, serologic monitoring of all cattle conducted via the same schedule as the vaccinations, which provided 6 cross-sectional blood sample collections of the 72 herds. Each blood sample was tested with an ELISA gE antibody test kit (HerdCheck , IDEXX, USA). Inconclusive results were considered as positive de Wergifosse et al. (1997). Whenever possible, the serologic status of calves that had positive results because of maternal antibodies was adjusted to that of the first test value obtained after those calves were 6 months old. In all other cases, cattle that had at least 1 positive test result were definitively considered as latent carriers and positive cattle for subsequent measurements. Calves < 6 months old for which no subsequent sample was available were excluded because there was no further analysis to define their immune status. New cases of BHV-1 infection were defined as gE seroconversion during the interval since the preceding sampling period.

Questionnaire for the risk factor analysis

Each owner completed a questionnaire about the herd management practices potentially linked to the risk of BHV 1 infection. The questionnaire was completed during a face-to-face interview. The questionnaire was designed on the basis of results for studies reported by Van Schaik et al. (1998) and Wentinck et al. (1993) of within-herd and between-herds risk factors and considered potential risk factors linked with herd management (ie, regular purchase of cattle, participation at fairs and shows, calf-dam relationship (whether calf was removed immediately after birth), use of artificial insemination or natural mating with bulls, and external sources of infection (ie, the presence of other farms within a radius of 100 m and possible contacts with other herds during the pasture period).

Research objectives of the IBR dataset

The conclusion of the review of the other studies is that there are no studies that compare the long term effect of different vaccination protocols on the hazard of BoHV-1 seroconversion. Most studies focus on clinical signs or virus titers after experimental challenge or reactivation but not under field conditions. Sometimes different protocols are compared using the basic reproduction number (R_0), the expected number of new cases of an infection caused by the introduction of an infected individual in a population of susceptible animals only. However, the concept of R_0 is very artificial and is only a summary statistic influenced by many unknown factors. There are also many different ways to estimate it, all with their own focus and corrections which make it difficult to compare the calculated R_0 from different studies (Heesterbeek, 2002). The long term effect on the incidence after implementation of a new vaccination protocol compared to the commonly used vaccination protocol is of major interest. A survival model gives more information about the evolution over time than the basic reproduction number. Variables can also be included in the survival model and their significance can be tested in a formal way.

The used control groups in the articles were placebo groups. In practice, most herds are vaccinated, using a less strict vaccination protocol and not a hyperimmunisation protocol. Therefore the use of a placebo groups was unethical in this setting (Temple and Ellenberg, 2000). Comparing a vaccine group with a non-vaccinated group (placebo) will result in a bigger effect than what the real effect would be in the field, where most farmers use a common vaccination protocol.

Analysis of experiments in the field are also complex due to the variability between farms, and the variability between seasons and even years, as suggested by the studies of Makoschey et al. (2007) and Nardelli et al. (2008). None of the mentioned studies took into account the fluctuations due to the differences between the seasons and years.

The aim of the analysis of the IBR dataset is to take all above mentioned problems into account using an appropriate survival model. First of all, a nonintervention group was used instead of a placebo group. The main interest of this study was to study the additional effect of hyperimmunisation compared to the common vaccination protocols. The nonintervention group did not change their common vaccination scheme, in most cases yearly booster vaccinations. The reason for this choice was twofold: it is unethical to force farmers to stop vaccinating against BoHV-1 and on the other hand, the aim of the study was to improve the current vaccination scheme, so a reduction of the hazard of seroconversion with the new hyperimunization protocols compared to the conventional protocol (nonintervention group) was the objective.

Secondly, most studies reduce the available longitudinal data (several samples on the same cow during the study period) to one single outcome: seroconverted at the end of the study or not. In this simplification the information of the time to seroconversion is lost. In our analysis, the time to seroconversion will be used as outcome variable, and since some cows did not seroconvert before the end of the study period, some times are right censored. Time to event (seroconversion) data with censoring are typically analyzed using survival models.

Only one study corrected for clustering, not by using a random effect but a fixed effect in the model. In the IBR dataset, cows are clustered within farms, they share the same environmental factors, and there are obviously differences between the farms. This is solved by introducing a shared gamma frailty as a random effect for farm.

The last problem was the seasonal effect. None of the mentioned studies corrected for the seasonal differences or differences between years although some authors suggest their existence (Makoschey et al., 2007; Nardelli et al., 2008). In the proposed model, the baseline hazard changes according to season. The estimated effects are the effects compared to the baseline hazard, which is thus corrected for seasonal and yearly fluctuations.

In the analysis, the effect of an HIG 1 and 2, compared to the reference group, NIG, on the hazard to have seroconversion against gE is modeled. When a HIG has a more protective effect than the common vaccination protocol, used in the NIG, it will have a negative effect over time on the hazard

to seroconvert (negative β means lower hazard, less chance to seroconvert).

1.3 The intramammary infection dataset

1.3.1 An introduction to intramammary infections and mastitis

Bovine mastitis, an inflammation of the mammary gland, is a complex disease resulting from interactions between the cow, microorganisms and the environment. It has a negative effect on the milk yield and milk quality (increase of somatic cell count, SCC). The economic impact results from the control costs (i.e. extra resource use, treatment,...) and losses due to reduced production. Therefore, in a lot of countries in the early sixties, programs were started to reduce the incidence of mastitis (Seegers et al., 2003; Neave et al., 1969).

The involved infectious organisms can be very diverse: bacteria, mycoplasma, yeasts and algae. Only bacteria are further considered in the dataset used in this thesis. Watts (1988) gave an overview of 137 different microbial species, subspecies and serovars isolated from the bovine mammary gland. Bradley (2002) gave an overview of the historical change in clinical mastitis incidence and its causes after the implementation of a mastitis control strategy, the Five-Point Plan. The plan has been very successful in reducing the incidence of clinical and subclinical mastitis due to contagious pathogens. Some pathogens (eg. *Streptococcus agalactiae*) are almost eradicated due to better sanitary measures. On the other hand, an increase of the incidence of environmental pathogens (mainly *S. aureus*) has been observed during the last years. One of the possible explanations is the reduced prevalence of the 'minor pathogens' (e.g. *Corynebacterium spp*) and, therefore, the removal of the protective effect that is often claimed in literature.

The role of viral infections in bovine mastitis is not very well documented until now. Some viruses have been isolated from milk from cows with (sub)clinical mastitis and some were detected in mammary tissue from cows with (sub)clinical mastitis. Some viruses may play a direct or indirect role in the etiology of mastitis due to their immunosuppressive properties (e.g. BoHV-1, BVD) or due to teat lesions (e.g. vesicular stomatitis, foot-and-mouth disease virus, bovine papillomavirus). Whether these viruses are able to induce bovine mastitis has not been reported (Wellenberg et al., 2002).

1.3.2 Pathogenesis of mastitis

The first step is the invasion of bacteria through the teat canal (due to contamination of the teat end) and the cistern. Most contagious pathogens, like *S. aureus*, adhere to the mammary epithelial cells and extracellular matrix components. They also invade in the mammary cells (mainly epithelial cells but also endothelial cells and fibroblasts). Most strains produce cell surface-associated and extracellular secretory product. *S. aureus* produces exotoxins that destroy tissue and protect bacteria against the host immune response (e.g. haemolytic toxins, α, β-toxins, enterotoxins, etc.) (Dego et al., 2002).

The invading bacteria interact with the mammary tissue and will induce an immune response, activating neutrophil migration from the blood to the alveoli and cisterns. This migration results in an often high increase of the SCC in the milk. Bacterial toxins, enzymes, and cell-wall components may have a direct effect on the epithelium of the mammary gland, but also stimulate the production of numerous mediators of inflammation. Infections with CNS and *C. bovis*, often called 'minor pathogens' result in a moderate inflammation with an increasing SCC, only two- to threefold the normal value. This results in a subclinical infection without visible changes of the milk composition and a slightly reduced milk production. Clinical mastitis is characterized by a swelling and pain in the udder, sometimes accompanied with systemic symptoms (anorexia, elevated body temperature, lethargy) and an abnormal milk composition. The magnitude of the inflammatory response is influenced by a lot of different factors such as the causative pathogen, age, stage of lactation, immune status, nutritional status and genetic factors (Harmon, 1994; Dego et al., 2002).

The initial inflammatory response consists of the influx of polymorphonuclear neutrophil (PMN) leukocytes into the mammary tissue. They appear in large numbers on the epithelium of the alveoli and in the lumen of the alveoli. The PMN engulf and digest the invading bacteria, but when they are in the lumen, they also engulf fat globules and casein which decreases their efficiency. They also release substances that attract more leukocytes to fight against the pathogen. In the case of persistent infections, the number of leukocytes can fluctuate but remains high, even some time after the bacteria were eliminated (Harmon, 1994).

The infection may cause also some other events. The bacterial toxins can damage the milk producing tissue which results in a reduction of the milk production. Also the PMN can harm the mammary tissue by releasing reactive oxygen intermediates and proteolytic enzymes. Leakage of blood

components into the milk can lead to aggregation of leukocytes and blood clotting factors. The formation of clots may obstruct the small ducts and block milk removal. In some cases this leads to a permanent loss of the function of that part of the gland, but functional tissue repair within the same or next lactation is also possible (Harmon, 1994; Zhao and Lacasse, 2008).

The last phase of the immune response during mastitis is the antibody based response. This has been studied most after infection with *E. coli* and *S. aureus*. Commercial vaccines are available for both, but their efficacy to protect against IMI is still debated (Schukken et al., 2011).

1.3.3 Mastitis causing pathogens

In the next section the most common mastitis causing pathogens in dairy cattle will be discussed. Most authors classify mastitis pathogens as either contagious or environmental, but this classification is not clear cut and is under discussion. Contagious pathogens are micro organisms adapted to survive within the mammary gland. Intramammary infection can result in clinical mastitis or a subclinical infection, which is characterized by an elevated somatic cell count (leukocytes and epithelial cells) in the milk of infected quarters, without the typical clinical symptoms. Cows are typically infected during milking from cow-to-cow. Only a few strains on each farm are responsible for IMI. The most common contagious pathogens are *S. aureus*, *S. dysgalactiae* and *S. agalactiae*.

Environmental pathogens are typically not adapted to survive within the mammary gland. They are opportunistic invaders, they multiply, induce a host immune response and are eliminated after a short time. Cow-to-cow spread is less common and a lot of different strains are observed in one farm. The most common environmental pathogens are the Enterobacteriaceae (particularly *E. coli*) and *S. uberis*. In literature a lot of different definitions are used to assign a pathogen to the group of contagious or environmental pathogens. A clear cut classification seems to be not feasible and gradual characterization is probably a better choice because some bacteria have both characteristics of environmental and contagious pathogens.

The last two discussed in this section are *C. bovis* and the coagulase-negative staphylococci (CNS), both minor pathogens commonly known to result in a subclinical mastitis. CNS is a very diverse group of different species with one common property: they are coagulase negative. In Chapter 4 the often claimed protective effect of those minor pathogens against other mastitis pathogens is investigated.

Staphylococcus aureus

S. aureus is probably the most studied mastitis pathogen in dairy cattle. There are many different strains within and between herds, but in most herds there is a single predominant strain which affects multiple cows. *S. aureus* enters first the teat orifice and can persist and multiply there and enters the teat canal by progressive colonization or by changes in intramammary pressure during milking. Colonization of the mammary gland may be achieved by adhesion to specific receptors on the surface of the epithelial cells (Dego et al., 2002).

The transmission is mainly cow-to-cow during milking (e.g. hands or towels contaminated by milk of an infected cow). Also the transmission by flies is suggested. Although most contagious mastitis pathogens are under control nowadays, the prevention of *S. aureus* mastitis seems to be difficult in some farms. A possible explanation is the poor response to treatment and the high number of false negative bacteriological cultures, resulting in undetected cases which may re-infect other cows. There is a high number of different strains and some are mainly found in milk or body sites while others are predominantly found in either milk or skin. The strain distribution is herd specific which makes it difficult to compare experiments performed in different farms (Zadoks et al., 2011).

Dufour et al. (2012) looked for additional manageable risk factors for the lactational incidence of *S. aureus*. The properties of the milking process were the most important risk factors. Beneficial sanitary measures were: adequate teat-end condition, disinfection and wearing gloves during milking which tends to prevent colonization of milkers' hands with transient flora such as *S. aureus*.

Streptococcus dysgalactiae

There is some discussion about the real nature of *S. dysgalactiae*. Some authors describe it as a contagious pathogen while others as an environmental pathogen, although environmental sources have not been investigated. But some properties are typical for environmental while others are typical for contagious pathogens. The infection is either transient or perisistent. There is in most cases one dominating strain within the herd and within one cow (Zadoks et al., 2011).

Streptococcus agalactiae

In cattle, mastitis is the only disease associated with a *S. agalactiae* infection. It is a typical contagious infection and its spread within the herd is mainly due to insufficient hygiene in the milking parlor and during milking (e.g. hands or towels contaminated by milk of an infected cow). Due to the strict cow-to-cow spread, there is often only one single strain present in a farm. This means that *S. agalactiae* is mainly a farm problem and can be eradicated by implementing sanitary measures and appropriate treatment programs. *S. agalactiae* was a big problem before the mastitis control programs were started but nowadays its prevalence is low (Zadoks et al., 2011). The efficacy of therapy on individual cows remains high and protocols for therapy of all infected animals in herd result generally in an eradication on the farm (Keefe, 1997).

***Escherichia coli* and other Enterobacteriaceae**

E. coli and the other Enterobacteriaceae have similar pathogenic properties and will be discussed together. *E. coli* is classified as an opportunistic environmental pathogen and an intramammary infection usually results in clinical mastitis, especially during parturition and early lactation. The production of lipopolysaccharide, an endotoxin, results in dose-dependent metabolic and clinical signs and is the main mediator for mastitis. A critical step in the defense against *E. coli* is the ability of the neutrophils to sequester and kill the bacteria. This can be influenced by hormones and metabolism of the cow. During the periparturient period, the innate immune system is compromised in many cows. The immunosuppression is the result of several physiologic changes during the transition period (Burvenich et al., 2007).

The isolates of coliform mastitis belong to a very large number of serological groups, mostly similar to the faecal isolates. Attachment to the mammary epithelium is not necessary and the type of *E. coli* strain is not the most important factor in determining the severity of the clinical mastitis (mild to fatal). Most studies show that adhesion is absent or very weak and there is no invasion. The severity of the mastis is mainly attributed to host characteristics (parturition, lactation stage, age, parity, metabolism, pro- and anti-inflammatory endogenous substances, growth factors, platelet activating factor, etc.) and management factors (e.g. stock density, nutrition). Most cases of *E. coli* mastis are transient and end with either death or full cure. Clinical mastitis with severe systemic symptoms occurs

at parturition and in early lactation. Infection during mid and late lactation results in mild to moderate mastitis. Recurrent mastitis cases are mainly due to repeated infections (with different strains) followed by cure. Persisting infections (mainly one single strain) with alternating mild clinical and subclinical episodes in the same quarter are also described (Burvenich et al., 2003; Hogan and Smith, 2003).

Klebsiella spp. is another common coliform of increasing importance on well-managed dairy farms. An IMI with *Klebsiella* spp. results in a more severe mastitis with a longer duration and higher production loss than *E. coli* (Schukken et al., 2012).

Streptococcus uberis

An intramammary infection with *S. uberis* may range from severe clinical mastitis to asymptomatic infection. *S. uberis* is usually classified as an environmental pathogen. The reservoir of infection is the environment of the bovine udder (body, manure, pasture and bedding materials) but also cow-to-cow transmission during the milking process is possible (Zadoks et al., 2003).

It is unclear if cow-factors or strain characteristics determine the duration of the infection and the severity of the mastitis. Some strains are predominantly associated with clinical mastitis while others with subclinical mastitis. Chronic subclinical infections with *S. uberis*, often unnoticed, are well known and can act as a reservoir for the rest of herd. The infection can persist for a long time but most infections are of short duration. There is also within-cow transmission, The infection is usually caused by the same strain when multiple quarters are infected (Zadoks et al., 2003).

Corynebacterium bovis

C. bovis causes mainly persistent IMI. They have a limited pathogenic potential and cause mainly subclinical mastitis, but sometimes a clinical mastitis is observed (Supré et al., 2011). *C. bovis* probably only colonizes the teat canal region (Pankey et al., 1985).

Coagulase-negative staphylococci

Coagulase-negative staphylococci (CNS) are nowadays the most common group of isolated bacteria from bovine milk in most western countries. CNS mainly originate from environmental reservoirs (body sites, farm environment) but also infected mammary glands can act as a source of infection.

CNS is a heterogeneous group of bacteria and the most common species are *Staphylococcus haemolyticus*, *Staphylococcus Simulans*, *Staphylococcus chromogenes* and *Staphylococcus epidermidis*. The last two did not seem to have a significant reservoir in the environment and seem to be of contagious nature with only a few predominant genotypes. *S. haemolyticus* and *S. simulans* are mainly found in the environment with different genotypes in the same farm (Piessens et al., 2012).

Almost all CNS species have very limited pathogenic potential and cause subclinical mastitis with a persistent IMI. Species differ substantially in their pathogenicity. Some species are mainly persistent infections (e.g. *S. chromogenes*) while others are mainly transient IMI (e.g. *S. haemolyticus* and *S. simulans*) (Supré et al., 2011).

1.3.4 Diagnosis of mastitis

Mastitis is a term used for both subclinical mastitis and clinical mastitis. The "golden standard" for masitis and IMI detection is bacteriological isolation. The IMI dataset is based on bacteriological sampling and more details are given in section 1.3.7. The rest of this section is about the alternatives for bacteriological isolation, which is expensive, time consuming and not always feasible for routine testing. Therefore alternative tests were developed for routine testing for subclinical mastitis, mainly based on the affected composition of the milk due to the inflammatory response (Pyörälä, 2003).

The milk somatic cells are mainly leukocytes (PMN, lymphocytes, macrophages), epithelial cells are less frequent. An increased SCC is not the result of epithelial cells but of an increased number of leukocytes due to inflammation (Harmon, 1994). Milk SCC has been used for a long time as indicator of IMI. The original limit for a healthy quarter was 500 000 cells/ml. Nowadays a limit of 100 000 cells/ml is commonly used in western countries. Quarters infected with major pathogens result in a SCC of more than 350 000 cells/ml. Coliform and *S. uberis* mastitis result frequently in SCC higher than a million cells/ml. The most used technique to determine the SCC is the electro-optical Fossomatic method (Pyörälä, 2003).

The California Mastitis Test (CMT, with a score ranging from 0 to 3) is a rapid technique to estimate the SCC, based on the DNA content of the milk. It is found to detect 75 to 80 % of the cows which needed therapy. A score 0 seems to correspond with a SCC of less than 200 000 cells/ml. The main advantages of the CMT are its low cost and the real-time results on the farm (Pyörälä, 2003).

There are many different tests available based on increased indigenous

enzymes in milk during mastitis. They include N-acetyl-β-D-glucosaminidase (NAGase), beta-glucuronidase and catalase. They are mainly experimental and the discriminatory capacity of these tests is rather low. These test are not (yet) suitable for large scale use (Pyörälä, 2003).

Measuring the electrical conductivity (EC) of milk to detect mastitis is based on the ionic changes in the milk due to the inflammation. Sodium and chloride concentrations increase. Electrical conductivity is highly correlated with the SCC. The EC measurement is converted to a computer readable signal and can be used as an on-line automatic monitoring system attached to the milking machine. Unfortunately also other, non-mastitis related, factors influence the ionic content of milk which reduces the diagnostic value of EC. This leads to a low sensitivity (less than 60%) and low specificity (around 70%) which make the predictive value of the method rather poor (Pyörälä, 2003).

1.3.5 Analysis of intramammary infections studies

A review of 13 studies that investigated the number of IMI is given in the next section. Most studies were longitudinal studies, with samples at predefined intervals, ranging from 2 to 4 weeks (Zadoks et al., 2001, 2002, 2003; Sommerhäuser et al., 2003; Parker et al., 2008) or only with samples when clinical mastitis was observed (Schukken et al., 1991; Lam et al., 1993; Barkema et al., 1999; Bradley and Green, 2001; Ferguson et al., 2007; Breen et al., 2009; Schwarz et al., 2010). The restriction of sampling only on clinical mastitis occasions is mainly due to financial reasons, bacteriological culture is an expensive analysis. This means that all above mentioned studies collected longitudinal data with a time to event nature. Moret-Stalder et al. (2009) took only samples on 2 occasions 2 weeks apart which is not longitudinal.

The sampling unit for bacteriology was in all cases quarter but Ferguson et al. (2007) also took sometimes composite samples of all functioning quarters. The occurrence of IMI was most often modeled using Poisson regression with the number of new infections as response variable (Schukken et al., 1991; Barkema et al., 1999; Zadoks et al., 2001; Sommerhäuser et al., 2003). Lam et al. (1993) only reported descriptive results about the incidence of IMI and Bradley and Green (2001) only performed a χ^2 test on the results. Zadoks et al. (2002) used a SIR-model (Susceptible, Infected, Recovered) to estimate the transition parameters from susceptible/uninfected to infected and from recovered to infected. In more recent papers logistic regression is often used to model IMI (Ferguson et al., 2007; Parker et al.,

2008; Breen et al., 2009; Moret-Stalder et al., 2009; Schwarz et al., 2010). None of the above mentioned analysis techniques took into account the time to event nature of the gathered data. Also loss to follow-up was not discussed in most papers, only Parker et al. (2008) discussed this problem in detail in his study.

Gasqui et al. (2000) introduced survival analysis as a tool to model clinical mastitis data (not IMI). A piecewise constant hazard model was used to model the time intervals between two consecutive cases of clinical mastitis during the same lactation. Zadoks et al. (2002) used a log rank survival test to model IMI. In most (older) studies udder quarters were treated as independent entities although they are clustered within cow and cows are clustered within farm. Schukken et al. (1991), Barkema et al. (1999) and Sommerhäuser et al. (2003) analyzed the data on herd level to avoid clustering problems. The main disadvantage of this technique is that the original data is reduced to the number of IMI on farm level. All information about the longitudinal nature of the data and information on quarter and cow level are totally ignored in the analysis.

In case of logistic regression, clustering on farm level is sometimes solved by the introduction of farm as a fixed effect (Zadoks et al., 2001; Ferguson et al., 2007) or in more recent studies as a random effect (Parker et al., 2008; Moret-Stalder et al., 2009; Breen et al., 2009; Schwarz et al., 2010). Additional correction of clustering on cow-level was only considered by some authors (Parker et al., 2008; Moret-Stalder et al., 2009; Breen et al., 2009). Although most papers mention the problem of clustering and some correct for it, only two papers quantified the magnitude of the clustering effect and interpreted these values (Barkema et al., 1999; Parker et al., 2008).

The main conclusion of the review of the used methodology is that the analysis of the data of most studies was not optimal. The time to event nature of the data is often reduced to binary data and the correlation structure of the data is often ignored. Ignoring clustering can lead to unreliable statistical significance tests due to the underestimation of the variance and, as a result, the underestimation of the type I error. A frailty model takes into account the time to event nature of the data, and also quantifies the clustering within cow. The stratified version of the model will also take into account the clustering on farm level. Unfortunately, none of the mentioned studies considered the use of a frailty model. The possible problems due to the non-continuous sampling (without the exact event time, interval censoring) was also not discussed in the papers.

Table 1.2: Overview of the studies of the effect of an intramammary infection with a minor pathogen. Experimental studies are denoted by 'exp', observational studies by 'obs'. The sample size is the number of quarters included in the study

Reference	type	sample size	sample unit	correction clustering	longitudinal data	lost to follow-up	analysis	significant result
Pankey et al. (1985)	exp	n.a.	quarter	n.a.	yes	n.a.	no	n.a.
Matthews et al. (1990)	exp	40	quarter	no	yes	n.a.	ANOVA	yes
Nickerson and Boddie (1994)	exp	1747	quarter	no	yes	n.a.	t-test	yes
Schukken et al. (1999)	exp	540	quarter	no/yes	yes	10 cows excl.	χ^2/GEE	no
Brooks et al. (1983)	obs	n.a.	quarter	no	no	-	χ^2	no
Hogan et al. (1988)	obs	1596	quarter	no	yes	n.a.	χ^2	no
Matthews et al. (1991)	obs	1339	quarter	no	yes	n.a.	χ^2	yes
Davidson et al. (1992)	obs	336	quarter	no	yes	yes	χ^2	no
Lam et al. (1997)	obs	1574	quarter	case-control	yes	n.a.	χ^2/OR	yes

1.3.6 Analysis of studies about the effect of an intramammary infection with a minor pathogen

A review of the methodology of 9 published studies about the (protective) effect of an IMI with a minor pathogen (*C. bovis* and CNS) on the susceptibility to a new IMI with other mastitis pathogens is given in the next section. An overview of the most important properties is given in Table 1.2.

The first group of studies are four experimental studies (Pankey et al., 1985; Matthews et al., 1990; Nickerson and Boddie, 1994; Schukken et al., 1999). The published study of Pankey et al. (1985) was unclear about the used methodology and no statistical analysis was performed and will not be further discussed due to a lack of information. The second group are five observational studies (Brooks et al., 1983; Hogan et al., 1988; Matthews et al., 1991; Davidson et al., 1992; Lam et al., 1997). The study performed by Brooks et al. (1983) was the only study were no longitudinal data were obtained and no sample size was provided. Also this study will not be further discussed in detail.

Except for the study of Brooks et al. (1983), all studies collected several samples at predefined intervals resulting in longitudinal data. For the analysis, none of the datasets were analyzed in a model that takes into account the longitudinal nature of the data. Most studies reduced the available time to event data to a binary variable: infected or not. This binary outcome variable was in most cases analyzed with a χ^2-test while Nickerson and Boddie (1994) used the t-test on the differences of the proportions of infected quarters (central limit theorem). Matthews et al. (1990) used the number of colony forming units as outcome variable and performed an ANOVA. The analysis of most studies are based on a chi-square statistic or regular logistic regression. None of the mentioned studies take into account the time to event nature of the original data due to the data reduction before analysis.

All studies sampled on quarter level. Correction for clustering of quarters within cow was only considered in 2 papers. Schukken et al. (1999) performed a general estimation equation logistic regression model that took into account the correlation within cow and an estimate of the intra class correlation was calculated (= 0.241) and indicated an important clustering. Lam et al. (1997) used a within-cow matched case-control approach to correct for possible confounding by herd, season and cow effects. All other studies did not control for clustering on cow or farm level.

Another important property in the analysis is the adjustment for loss to follow-up. Loss to follow-up was only mentioned in two papers. Schukken et al. (1999) excluded animals with missing values. Davidson et al. (1992) only

mentioned that 28 cows had missing samples but it is not clear if they are excluded in the analysis or not. It is very likely that also the other studies had missing values and excluded those quarters/cows for further analysis but this information is not stated in the articles.

The last problem is the time-varying nature of the infection status. In the observational studies it was often assumed that the infection status remained constant during the study period, although this is not always the case. Davidson et al. (1992) solved this issue by analyzing the eight sampling times separately which makes it difficult to interpret the results. Hogan et al. (1988) analysed the quarter-day exposure for uninfected and infected quarters to correct for this issue.

The main conclusion of the review of the used methodology is that the analysis of the data of all studies was not optimal. A frailty model takes into account the time to event nature of the data, the clustering within cow, time varying covariates and missing values due to censoring. Unfortunately, none of the mentioned studies considered the use of a frailty model.

1.3.7 The dataset

Design

In total, 1132 cows on 25 dairy herds, located in the provinces East and West Flanders, Belgium, were followed during a 20-months period (February 1993 to September 1994). Criteria for herd selection were willingness of the farmer to cooperate in the observational study, participation in the dairy herd improvement program (DHI), organized by the Flemish Cattle Breeding Association and a minimum herd size of 25 cows. Only the lactations of cows with parturition after February 1993 were considered for analysis.

Bacteriological monitoring

At monthly intervals (11 times a year, not during either July or August) quarter foremilk samples were taken to detect intramammary infections (IMI). The teats were cleaned with dry udder cloths. Dirty teats were washed and dried. Before milk samples were taken, all teats were disinfected with cotton moistened with asolution of ethyl alcohol (70%) and chlorhexidine(200 mg/100 ml). Vangroenweghe et al. (2001) showed that there is no significant difference in bacterial contamination between manual sampling and aseptic collection through a sterile cannula. Immediately after collection, the milk samples were transported to a laboratory and 0.01 mL aliquots were streaked for initial isolation within 2 to 3 h after collection onto

a 90-mm Petri dish with a blood agar base (Oxoid, Basingstoke, England) supplemented with 5% bovine blood. Samples were also streaked onto an Edwardsmedium (Oxoid) supplemented with 5% bovine blood. Agar plates were incubated at 37°C and read after 24 and 48 h (Laevens et al., 1997).

A quarter was considered to be positive when more than 100 cfu/mL of the considered pathogen was found. In literature a lot of different definitions are used to define the infection status. Dohoo et al. (2011a) gave an overview of the most common definitions and estimated the sensitivity and specificity of several definitions (single, duplicate and triplicate quarter milk samples, minimum colony count per mL). Triplicate milk samples (during 5 days) are often considered as the golden standard detection method for an IMI. Although triplicate and duplicate samples provided the best combination of specificity and sensitivity, only a modest gain in specificity and little or no gain in sensitivity was observed. This suggests that the additional expense of multiple samples is not justified and a single quarter milk sample is in most situations the best choice (Dohoo et al., 2011a). Also the number of detected colonies in a 0.01 mL aliquot has to be defined. In our dataset 1 colony (=100 cfu/mL) was considered as a positive sample . Dohoo et al. (2011b) suggest that this definition is an acceptable definition when identifying as many existing infections as possible is important. This results for most pathogens in a sensitivity ranging from 75% to 90% but a high specificity (over 97% except for CNS: 87%).

Different bacteria were isolated and identified as described by the National Mastitis Council. In this PhD we focus on *Staphylococcus aureus*, *Streptococcus uberis*, *Streptococcus dysgalactiae*, *Escherichia coli*, other enterobacteriaceae, *Corynebacterium bovis* and coagulase-negative staphylococci (CNS). The distribution of the number of new IMI's with the bacteria is given in Table 1.3. The other isolated bacteria were too uncommon and not considered for further analysis: *Streptococcus agalactiae*, esculin-positive cocci, corynebacteria excluding *C. bovis*, *Arcanobacterium pyogenes*, *Klebsiella spp.*, *Serratia spp.*, *Pseudomonas aeruginosa*, yeasts and fungi.

In the original dataset, the information of each sample was available on quarter level (see Table 1.4). In the example, cow 89407831117 of farm 1 was sampled 8 times (on day 40, 75, 96, 124, 156, 190, 239, 267). At day 40, *C. bovis* was found in all quarters. At day 75 in the left rear quarter *S. dysgalactiae* was found, the other quarters were negative for bacteriology. At day 96 all quarters were negative (quarter = n.a. and bacteriology = 0)

Table 1.3: Number of new infections with different bacteria in the IMI dataset.

	Number	% Quarter
Staphylococcus aureus	246	5.0
Streptococcus uberis	198	4.0
Streptococcus dysgalactiae	97	2.0
Escherichia coli	34	0.7
other enterobacteriaceae	54	0.1
Corynebacterium bovis	1976	40.0
Coagulase-negative staphylococci (CNS)	827	16.7

and the left front quarter became positive for *S. aureus* at day 124. This means that the infection with *S. aureus* occurred between 96 and 124 days in lactation (see Table 1.5). At day 190 coagulase-negative staphylococci were isolated from the right rear quarter.

Table 1.4: Structure of the original IMI dataset with the results of the bacteriological isolations.

Cow nr	lactationday	date	quarter	bacteriology
89407831117	40	10-Dec-93	LR	CBO
89407831117	40	10-Dec-93	LF	CBO
89407831117	40	10-Dec-93	RR	CBO
89407831117	40	10-Dec-93	RF	CBO
89407831117	75	14-Jan-94	LR	SDY
89407831117	96	04-Feb-94	n.a.	0
89407831117	124	04-Mar-94	LF	SAU
89407831117	156	05-Apr-94	RR	CBO
89407831117	156	05-Apr-94	RF	CBO
89407831117	190	09-May-94	LR	CBO
89407831117	190	09-May-94	RR	CNS
89407831117	239	27-June-94	RR	CBO
89407831117	239	27-June-94	RF	CBO
89407831117	267	25-July-94	n.a.	0
...				

The original data were rearranged in two different ways according to their use as either an outcome variable (time to event) or a time varying covariate. For the time to event dataset (Table 1.5) the time to infection with a specific pathogen was recorded. As a result of the monthly samples, the exact infection times were not known and the time of the last negative sample was recorded as the lower bound of the interval while the first positive sample was recorded as the upper bound of the interval. The event is only known to have occurred between the lower and upper bound. In Table 1.5, only the left front quarter of the primiparous (parity = 0) cow 89407831117 of farm 1 was infected (censor indicator = 1) between 96 (lower) and 124 (upper) days in lactation.

Table 1.5: Structure of the time IMI with *Staphylococcus aureus* dataset.

Farm nr	Cow nr	parity	quarter	lower	upper	censor
1	89407831117	0	LR	238	267	0
1	89407831117	0	LF	96	124	1
1	89407831117	0	RR	239	267	0
1	89407831117	0	RF	239	267	0
1	89409376144	1	LR	284	313	0
1	89409376144	1	LF	284	313	0
1	89409376144	1	RR	284	313	0
1	89409376144	1	RF	284	313	0
...						

In case of CNS and *C. bovis*, the infection status was used as a time varying covariate. In the infection status dataset, the monthly infection status of each quarter during the whole lactation was recorded. A quarter was assumed to be infected starting from the month before the observed isolation of CNS or *C. bovis*. As long as the germ was isolated during the following months, the quarter was assumed to be infected during those months. Dohoo et al. (2011) showed that culture procedures based on one milk sample, have a limited sensitivity, particularly in case of CNS and Streptococcus spp. Therefore, a negative sample preceded and followed by a positive sample was assumed to be a false negative value and converted to a positive value. In

Table 1.6 all quarters of cow 33333831142 remained uninfected until day 50. At day 50 both right quarters and the left front quarter became infected with CNS and at day 84 also the left rear quarter became infected. All quarters remained infected until the end of the follow-up period (for this cow 274 days). For cow 81343911121 all quarters remained uninfected until the end of the follow-up period (185 days).

Table 1.6: Structure of the time varying covariate, intramammary infected with coagulase-negative staphylococci or not (variable CNS) dataset. The variables 'start' and 'end' are the start time and the end time of the period where the infection status remained constant.

Farm nr	Cow nr	quarter	start	end	CNS
1	33333831142	LR	0	84	0
1	33333831142	LR	84	274	1
1	333338311421	LF	0	50	0
1	33333831142	LF	50	274	1
1	33333831142	RR	0	50	0
1	33333831142	RR	50	274	1
1	33333831142	RF	0	50	0
1	33333831142	RF	50	274	1
16	81343911121	LR	0	185	0
16	81343911121	LF	0	185	0
16	81343911121	RR	0	185	0
16	81343911121	RF	0	185	0
...					

Research objectives of the IMI dataset

In most previously mentioned studies (see section 1.3.5 and 1.3.6), cows are evaluated after a certain risk time with a binary outcome: infected or not infected (although infection times are often known). In most cases the analysis was performed on quarter level but only a few analyses corrected for clustering of quarters within a cow. Most other papers assume independence of the udder quarters despite the important role of cow factors in the occurrence of IMI and the severity of mastitis.

The clustering is of general interest. In Chapter 3 a new technique is proposed to analyze interval censored and clustered survival times in an analytical way. In the analysis, the effect of parity on the hazard to become infected with *C. bovis*, *S. uberis* and *S. aureus* is modeled.

Quarters are clustered within a cow, but cows are also clustered within a farm. To have an idea of the contribution of the clustering on cow and on farm level, a stratified (by farm) analysis is performed, on a subset of the data. In the stratified analysis, the clustering is only due to the cow factors. The variable of interest was parity, primiparous versus multiparous.

In Chapter 4, the technique explained in Chapter 3 is extended to include also time varying variables. All previously mentioned papers that investigated the effect of CNS or *C. bovis* on the incidence of other mastitis pathogens divided the study population in two groups: either infected or not infected with CNS during the lactation. The occurrence of IMI with major pathogens was compared between the infected and uninfected group. This assumes that the infection with CNS is persistent during the entire study period. However, in observational studies some quarters become infected, the infection is persistent for some time and the CNS are eliminated after some period. The infection status is therefore not constant over time but is time dependent.

A new technique is developed to include timevarying variables in the shared frailty model for interval censored data. In the analysis, the effect of an IMI with CNS or *C. bovis*, as a time varying covariate, on the hazard to have an IMI with *S. aureus*, *S. uberis*, *S. dysgalactiae*, *E. coli* and other coliform bacteria is modeled with correction for parity.

1.4 Survival data

Survival data, time-to-event data, lifetime, failure time data are all different names for the same data type: describing the time to an event. The event may be death (literally survival), but other events, infections, seroconversions, the onset of a disease are possible terms. Survival models are often used in biomedical statistics, but also in engineering, e.g. to model the expected lifetime of mechanical and electrical equipment (Widodo and Yang, 2011), in economics, e.g. to model the time to bankruptcy or credit risk (Luoma and Laitinen, 1991; Im et al., 2012), the time to sale of products (Brint, 2012), etc.

One of the main properties of survival data is the occurence of censoring. In a time-to-event study, data are typically collected during a predefined

period. At the end of the study, an event will have occured for some subjects while for other subjects no event was observed, which is right censored. For such subjects we only known that the event time falls after that time, this is called a right censored subject.

1.4.1 Basic notation and functions in survival analysis

Let $f(t)$ denote the density function of T, an absolutely continuous, non-negative random variable representing the time to event. Integrating the density function over time leads to the corresponding cumulative distribution function:

$$F(t) = P(T < t) = \int_0^t f(s)ds$$

The nonincreasing survival function ($S(t)$) is defined as the probability that the event time T exceeds a value t (Kaplan, 1958):

$$S(t) = 1 - F(t) = P(T \geqslant t) = \int_t^\infty f(s)ds.$$

An important concept in survival analysis is the hazard function $h(t)$, which represents the instantaneous failure rate at time t, given that the subject did not fail until time t. The hazard function is defined as

$$h(t) = \lim_{\Delta t \to 0^+} \frac{P(t \leqslant T < t + \Delta t | T \geqslant t)}{\Delta t}.$$

Integrating the hazard function over time leads to the corresponding cumulative hazard function:

$$H(t) = \int_0^t h(s)ds$$

The survival, density and hazard functions have following one-to-one relationships:

$$\begin{aligned} f(t) &= -\frac{dS(t)}{dt} \\ h(t) &= \frac{f(t)}{S(t)} = \frac{-d \log S(t)}{dt} \\ S(t) &= \exp(-H(t)) \end{aligned}$$

1.4.2 Censoring

Censoring is a specific feature of survival data. Throughout this thesis, noninformative censoring is assumed. This means that the censoring time is independent of the event time. This is an essential assumption to ensure that censoring does not have an impact on the parameter estimates.

Left, right and interval censoring are different types of censoring. Right-censoring occurs when the event time is larger than the followup time. Possible reasons for right censoring are the end of the study, loss to follow-up of subjects (drop out due to another disease, voluntarily exit from the study, dead due to another reason etc.).

Left censoring occurs when the event time is smaller than the censoring time, i.e., when the event already occurred before the subject entered the study. Interval-censored data arise when the exact event time is not known; it is only known that the true event time is contained in the interval between the last observed time without the event (lower bound) and the first time with the event (upper bound).

A response of a subject typically consists of two parts, a time indication and a consoring indicator, δ_i, taking the value 1 if the event has been observed, otherwise, in case of right censoring, δ_i takes the value 0. Different censor schemes, presented in Figure 1.1 and 1.2, lead to different types of response information.

In Figure 1.1 a study was conducted during a follow-up period of 210 days. If the exact time is known (subject 1), the response consists of that time (160d) and the censoring indicator is equal to 1, which indicates that it is an event time. If there was no event before the end of the study (subject 2), the time will be the end of the follow-up period (210d) and $\delta_i = 0$, which indicates that it is a right censored time. If a subject leaves the study before the end of the study (subject 3), the time will be the lost to follow-up time (90d) and $\delta_i = 0$.

In Figure 1.2 a study was conducted during a follow-up period of 210 days but the subjects were tested only each 30 days (on day 30, 60, 90...). Subject 1 had a positive test at day 150, the last negative test was on day 120. The unobserved event occured in the interval 120-150 days, the response consists of the lower and upper bound of the interval (120d, 150d) and the censorindicator is equal to 1, which indicates that it is an interval censored time. If there was no event before the end of the study (subject 2), the time will be the end of the follow-up period (210d) and $\delta_i = 0$, a right censored time. If a subject leaves the study before the end of the study (subject 3), the time wil be the last observation time before the subject left

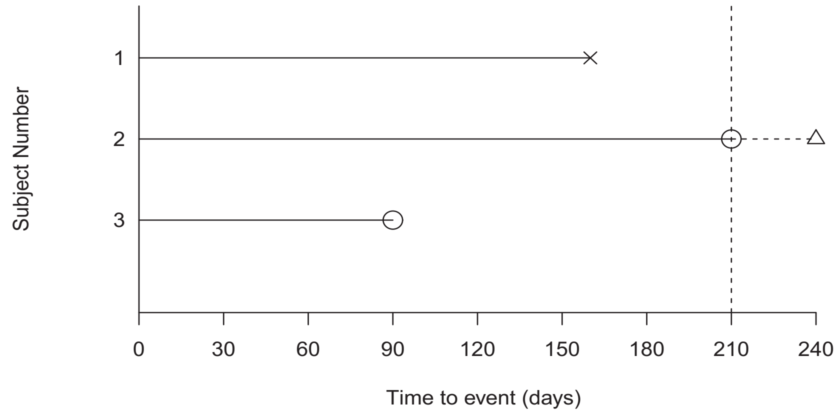

Figure 1.1: Exact (subject 1) and right censored (subjects 2 and 3) observations. The symbol 'x' denotes an observed event time, the symbol 'o' a censored observation time and the symbol '△' denotes an unobserved event time.

the study (120d) and $\delta_i = 0$, a right censored time.

Left- and right-censored observations can be considered as special cases of interval-censoring. For a right-censored observation, the start of the interval is the censoring time and the upper bound of the interval is infinity. When the event time is known exactly, the lower bound is equal to the upper bound.

In the case of right-censored data, there are either exact event times or right-censored observations. The likelihood for a sample of size n is given by (Klein and Moeschberger, 1997)

$$L = \prod_{i=1}^{n} \left(f(y_i)\right)^{\delta_i} \left(S(y_i)\right)^{1-\delta_i} .$$

where $f(.)$ the density function of the event times and $S(.)$ the cumulative distribution function of censoring times and δ_i the censoring indicator.

The contributions of survival data with different censoring mechanisms in case of noninformative censoring, are given by (Klein and Moeschberger, 1997):

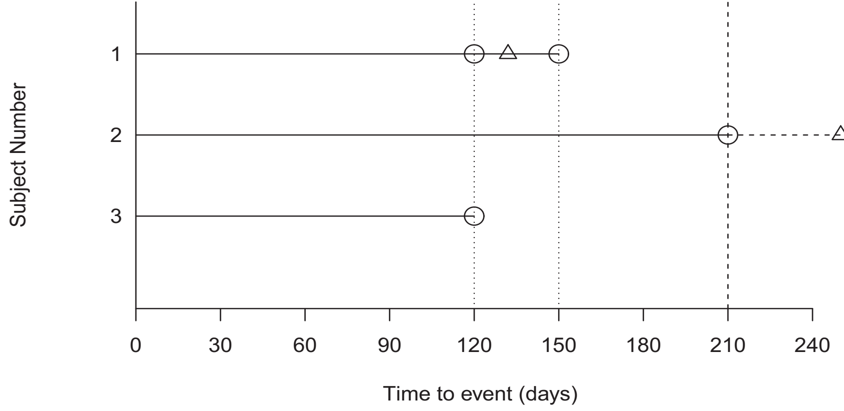

Figure 1.2: interval censored (subject 1) and right censored (subjects 2 and 3) observations. The symbol '△' denotes an unobserved event time, the symbol 'o' a censored observation time.

exact event times:	$f(y_i)$
right-censored observations:	$S(y_i)$
left-censored observations:	$1 - S(y_i)$
interval-censored observations:	$(S(L_i) - S(R_i))$

The likelihood function may be constructed by combining the different components. The most general form is:

$$L = \prod_{i \in D} f(y_i) \prod_{i \in R} S(y_i) \prod_{i \in L} (1 - S(y_i)) \prod_{i \in I} (S(L_i) - S(R_i))$$

with D the set of death times, R the set of right-censored observation, L the set of left-censored observations and I the set of interval-censored observations. In this thesis left censoring in not further considered. The likelihood of the IMI dataset, a combination of right censored and interval censored data, can be written as:

$$L = \prod_{i=1}^{n} \left(S(L_i) - S(U_i)\right)^{\delta_i} \left(S(R_i)\right)^{(1-\delta_i)} \tag{1.1}$$

where L_i and U_i the lower and upper bound of the intervals of the interval-censored observations and R_i the right censored time. δ_{ij} is the censoring indicator (0: censored, 1: event, interval censored).

1.4.3 The proportional hazards model

The most popular regression model for survival data, especially in the field of medicine and biostatistics, is the proportional hazards model (Cox, 1972). The hazard function $h_i(t)$ of subject i, is the product of a common unspecified baseline hazard function $h_0(t)$ and the exponential of a linear function of $\boldsymbol{x_i}$, the vector of covariates for subject i.

$$h_i(t) = h_0(t) \exp(\boldsymbol{x}_i^t \boldsymbol{\beta}),$$

with $\boldsymbol{\beta}$ the vector of the regression parameters. The common baseline hazard $h_0(t)$ is assumed to be the same for all subjects and can have a specific form (the parametric proportional hazards model) or can be unspecified (the semiparametric proportional hazards model).

The ratio of the hazard functions for two subjects with different covariate information is assumed to be constant over time (proportional hazards). Consider a simple example with a treatment group ($x_i = 1$) and a control group ($x_i = 0$). The hazard ratio is equal to:

$$HR = \frac{h_0(t) \exp \beta}{h_0(t)} = \exp \beta.$$

In the case of a semiparametric model, the baseline hazard $h_0(t)$ is left unspecified. The effect of the covariates on the hazard function is modeled parametrically, hence the term semiparametric. One of the main reasons for the popularity of the semiparametric Cox proportional hazards model is the existence of a simple and efficient inference procedure for the regression parameters in case of right-censored data, the partial likelihood maximization procedure introduced by Cox (1972). This likelihood is only a function of the unknown regression parameters $\boldsymbol{\beta}$ and does not contain the baseline hazard $h_0(t)$. The regression parameters can be estimated through maximization of the partial likelihood. The technique is also implemented in most statistical software packages (e.g. coxph() in S-plus and R; PROC PHREG in SAS).

In the parametric proportional hazards model, the baseline hazard function $h_0(t)$ is assumed to have a particular parametric form. A popular assumption for the parametric baseline hazard is the Weibull distribution:

$$h_0(t) = \lambda\gamma t^{\gamma-1},$$

with $\gamma > 0$ the shape parameter and $\lambda > 0$ the scale parameter. The Weibull distribution is a popular choice as it is a flexible distribution that often describes the evolution of the hazard well in practice. The shape parameter γ has an impact on the shape of the density curve. For $\gamma < 1$ the hazard decreases monotonically over time, for $\gamma > 1$ the hazard is monotone increasing and for $\gamma = 1$ the hazard is constant over time, corresponding to exponentially distributed event times. Figure 1.3a shows the effect of different scale parameters on the Weibull hazard function with a fixed shape parameter ($\gamma = 1.1$) while Figure 1.3b shows the effect of different shape parameters with a fixed scale parameter ($\lambda = 0.03$). Other possible choices for the distribution of the event times include the exponential, Gompertz, loglogistic and lognormal distribution (Marshall and Olkin, 2007).

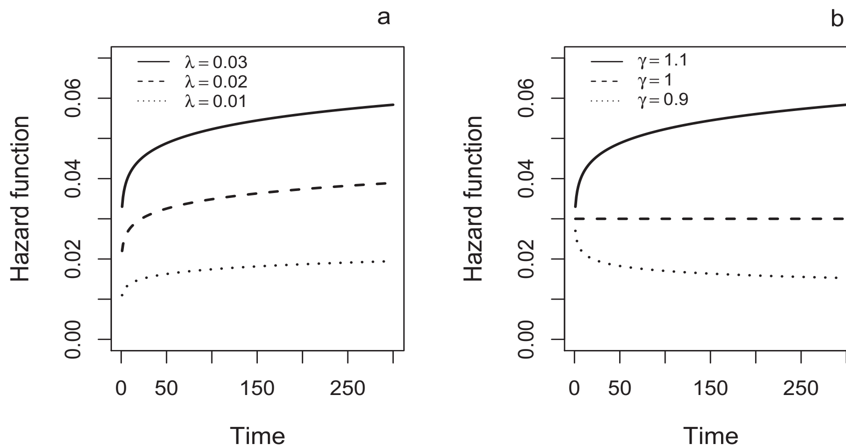

Figure 1.3: Weibull hazard functions with (a) different scale (fixed $\gamma = 1.1$) and (b) different shape parameters (fixed $\lambda = 0.03$).

1.4.4 Multivariate survival data: the frailty model

The techniques in previous sections assume that survival times of different subjects are independent. This assumption may be valid in some situations, but in veterinary science most survival times are not independent of each other because some subjects share the same environment. For example, different cows are housed in the same farm, different quarters belong to the same udder, different piglets originate from the same litter, etc. There are several techniques to correct for the correlation between the event times. One option is the marginal survival model based on robust variance estimation (Stedman et al., 2012). Another approach is the copula model where a copula function couples the marginal survival functions and the joint survival function and determines the type of correlation. Goethals et al. (2008) demonstrate the use of copulas as an alternative for the shared frailty model. In this section the focus is on the univariate shared frailty model which provides an estimate of the correlation between the event times in a cluster.

The term frailty was introduced by Vaupel et al. (1979) in a univariate frailty model, where an individual with a frailty of 1 might be called a "standard" individual while an individual with a frailty larger than 1 is more likely to have an event at any particular age. The frailty, a random quantity, is often assumed to be gamma distributed, because of its nice analytical properties, its strictly positive nature and its flexibility. The variance of the frailty distribution is a measure of the heterogeneity between the subjects in the study population: the higher the variance, the more heterogeneous the population.

In medical science, "frailty" was commonly used to describe an increased risk to several diseases, a higher tendency to functional failure or even death. It is used mostly to describe, in geriatrics, the last state before profound functional loss, almost always followed by death. A clear definition of frailty is difficult and there is no precise scientific meaning. The criteria to define that a person is frail, is a very complex and highly demanding task. The frailty of a person, the tendency to fail, is influenced by a complex interaction of a lot of known and unknown factors such as: genetic disorders, (infectious) diseases, injuries, lifestyle, nutrition, aging, etc. (Bortz, 2002).

The frailty in medical science is basicly at the indiviual level. In survival modelling, a frailty can be individual (the univariate frailty model) but in most cases subjects are characterized by the same frailty as they share the same environment. In the IBR dataset, all cows of the same farm share the same factors on that farm (e.g. management factors, number of buildings,

herdsize, but also a lot of unknown factors). The farm level can then be introduced in the model as a frailty term, and all cows in the same farm are assumed to share that same frailty. In this thesis, only the shared frailty model is used.

The proportional hazards model explained in Section 1.4.3 can be extended to include a frailty term. All subjects in the same cluster share the same frailty term, hence the name shared frailty model. The shared frailty u_i is common to all subjects in one cluster, which creates correlation between the event times of the subjects in the cluster.

We introduce the model using the structure of the IMI dataset. We have 4 quarters in each cow (=cluster) in a sample with k cows. The shared frailty model for this example is given by:

$$h_{ij}(t) = h_0(t) u_i \exp(\boldsymbol{x}_{ij}^t \boldsymbol{\beta}),$$

where $h_{ij}(t)$ is the conditional hazard function (conditional on u_i) at time t for the jth quarter of the ith cow, $j = 1, 2, 3, 4$, $i = 1, \dots k$, $h_0(t)$ is the baseline hazard, $\boldsymbol{x}_{ij}$ is the vector of covariates for the jth quarter of cow i, $\boldsymbol{\beta}$ is the vector of regression parameters and u_i is the frailty for cow i.

The frailties (u_i) are assumed to come from a density $f_U(u)$. There are several possible choices for densities of the frailties including the gamma distribution, the inverse Gaussian distribution, the positive stable distribution, the power variance function, the compound Poisson distribution and the lognormal distribution (Duchateau and Janssen, 2008). The most common distribution is the one-parameter gamma distribution because of its nice mathematical properties. We futher consider only the one-parameter gamma distribution (gamma$(1/\theta, 1/\theta)$) with density:

$$f_U(u) = \frac{u^{1/\theta - 1} \exp(-u/\theta)}{\theta^{1/\theta} \Gamma(1/\theta)}, \tag{1.2}$$

with $\theta > 0$ and $\Gamma(.)$ the gamma function.

The distribution has mean 1 and variance θ. Beside its nice mathematical properties, there are no biological reasons that justify the choice of a gamma distribution. However, in practice, misspecification of the frailty distribution results generally in low bias for the parameter estimates. This suggests that the gamma distribution is a good choice in practice, especially when the regression parameters are of primary interest (Hsu et al., 2007).

In case of right censoring and a gamma distribution for the frailties, the

conditional likelihood for cow i is:

$$L_i(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\beta}|u_i) = \prod_{j=1}^{n_i} (h_0(y_{ij})u_i \exp(\boldsymbol{x}_{ij}^t\boldsymbol{\beta}))^{\delta_{ij}} \exp(-H_0(y_{ij})u_i \exp(\boldsymbol{x}_{ij}^t\boldsymbol{\beta}))$$

with n_i the number of subjects in cluster i (in the example there are 4 quarters in each cow, $n_i = 4$, $i = 1, ..., k$), $\boldsymbol{\xi}$ containing the parameters of the baseline hazard (in case of Weibull: λ and γ). The marginal likelihood for the ith cluster is:

$$\begin{aligned} L_{marg,i}(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\beta}) &= \int_0^\infty \prod_{j=1}^{n_i} (h_0(y_{ij})u \exp(\boldsymbol{x}_{ij}^t\boldsymbol{\beta}))^{\delta_{ij}} \exp(-H_0(y_{ij})u \exp(\boldsymbol{x}_{ij}^t\boldsymbol{\beta})) \\ &\quad \times \frac{u^{1/\theta-1}}{\theta^{1/\theta}\Gamma(1/\theta)} \exp\left(-u/\theta\right) du. \end{aligned}$$

The frailties can be integrated out from the conditional likelihood in an analytical way, resulting in a simple and closed form expression for the marginal likelihood. The marginal likelihood for cow i is then:

$$L_{marg,i}(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\beta}) = \frac{\Gamma(d_i + 1/\theta) \prod_{j=1}^{n_i} (h_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^t\boldsymbol{\beta}))^{\delta_{ij}}}{\left(1/\theta + \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^t\boldsymbol{\beta})\right)^{1/\theta + d_i} \theta^{1/\theta}\Gamma(1/\theta)}$$

with $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ the number of observed events in cluster i.

The marginal loglikelihood is then given by (Klein and Moeschberger, 1997):

$$\begin{aligned} l_{marg}(\boldsymbol{\xi}, \boldsymbol{\theta}, \boldsymbol{\beta}) &= \sum_{i=1}^{k} \Bigg[d_i \log\theta - \log\Gamma(1/\theta) + \log\Gamma(1/\theta + d_i) \\ &\quad -(1/\theta + d_i) \log\left(1 + \theta \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(\boldsymbol{x}_{ij}^t\boldsymbol{\beta})\right) \\ &\quad + \sum_{j=1}^{n_i} \delta_{ij} \left(\boldsymbol{x}_{ij}^t\boldsymbol{\beta} + \log h_0(y_{ij})\right) \Bigg], \end{aligned}$$

If a parametric assumption is made for the baseline hazard (e.g. Weibull), the marginal loglikelihood is fully parametric and classical maximisation techniques can be used (e.g. nlmin in S-plus, nlm in R) to obtain parameter estimates. Standard errors can be obtained from the inverse of the observed information matrix. More details about the maximization are given by Duchateau and Janssen (2008). This basic model is further extended to interval censoring in Chapter 3 and time-varying covariates are added in Chapter 4

1.5 Research objectives

The main objective of this dissertation is to introduce the shared frailty model methodology to model the dynamics of infectious animal diseases. Modeling the dynamics of infectious animal diseases makes it possible to improve the understanding of the disease dynamics, identify the risk factors and evaluate the control measures.

In evaluating vaccination protocols in IBR, the hazard of seroconversion needs to be modeled in terms of calendar time, taking into consideration as well the the seasonal variations in the infection pressure and clustering of animals within a farm. A piecewise constant hazard shared frailty model was used to analyze the data introducing a shared frailty on farm level to account for the clustering. The evolution over time, since the start of the hyperimmunization protocol on the farm, was modeled to estimate the long term effect of the hyperimmunization. The follow-up time was divided in summer and winter periods, to allow the baseline hazard to change according to the season. The model and the results are explained in Chapter 2.

The IMI data are characterized by interval censoring and clustering. Due to the monthly samples, the true infection time is contained in an interval between the last observed time without infection and the first time with infection. A new technique for interval-censored data, where the frailties are integrated out in an analytical way, is developed in Chapter 3 and used to model the effect of parity on the hazard to have an IMI with *C. bovis*, *S. uberis* and *S. aureus.* In another analysis, stratified by herd, the importance of clustering on farm level and cow level is estimated.

In Chapter 4, the technique explained in Chapter 3 is extended to include also time varying covariates in order to model the effect of an IMI with CNS or *C. bovis* on the susceptibility to a new IMI with other mastitis pathogens. The time to IMI with a mastitis pathogen was determined for

each udder quarter and constitutes the interval censored response variable. The infection status of the quarter with CNS or *C. bovis*, allowed to change each month, is a time varying covariate. Parity is also included as a fixed effect.

Bibliography

Babiuk L.A., Littel-van den Hurk, S.V. and Tikoo S.K., 1996. Immunology of bovine herpesvirus 1 infection. Vet. Microbiol. 53, 31-42.

Barkema H., Bartels C., Van Wuijckhuise L., Hesselinck J., Holzhauer M., Weber M., Franken P., Kock P., Bruschke C. and Zimmer G., 2001. Outbreak of bovine virus diarrhea on Dutch dairy farms induced by a bovine herpesvirus 1 marker vaccine contamined with bovine viral diarrhea virus type 2. Tijdschr. Diergeneeskd. 126, 158-165.

Barkema H.W., Schukken Y.H., Lam T.J.G.M., Beiboer M.L., Benedictus G. and Brand A., 1999. Management practices associated with the incidence rate of clinical mastitis. J. Dairy Sci. 82, 1643-1354.

Boelaert F., Biront P., Soumare B., Dispas M., Vanopdenbosch E., Vermeersch J.P., Raskin A., Dufey J., Berkvens D. and Kerkhofs P., 2000. Prevalence of bovine herpesvirus-1 in the Belgian cattle population. Prev. Vet. Med. 45, 285-295.

Bortz W.M., 2002. A conceptual framework of frailty: A review. J. Gerontol. Ser. A-Biol. Sci. Med. Sci. 57, M283-M288.

Bosch J.C., De Jong M., Franken P., Frankena K., Hage J., Kaashoek M.J., Maris-Veldhuis M., Noordhuizen J., Van Der Poel W., Verhoef J., Weerdmeester K., Zimmer G. and Van Oirschot J.T., 1998. An inactivated gE-negative marker vaccine and an experimental gD-subunit vaccine reduce the incidence of bovine herpesvirus 1 infections in the field. Vaccine 16, 265-271.

Bosch J.C., Kaashoek M.J., Kroese A.H. and van Oirschot J.T., 1996. An attenuated bovine herpesvirus 1 marker vaccine induces a better protection than two inactivated marker vaccines. Vet. Microbiol. 52, 223-234.

Bosch J.C., Kaashoek M.J. and van Oirschot J.T., 2003. Inactivated bovine herpesvirus 1 marker vaccines are more efficacious in reducing virus excretion after reactivation than a live marker vaccine. Vaccine 15, 1512-1547.

Bradley A.J., 2002. Bovine mastitis: an evolving disease. Vet. J. 164, 116-128.

Bradley A.J. and Green M.J., 2001. Aetiology of clinical mastitis in six Somerset dairy herds. Vet. Rec. 148, 683-686.

Breen J.E., Bradley A.J. and Green M.J., 2009. Quarter and cow risk factors associated with a somatic cell count greater than 199,000 cells per milliliter in United Kingdom dairy cows. J. Dairy Sci. 92, 3106-3115.

Brint A., 2012. Modelling a property's time to sale by inflating its previous selling price. J. Oper. Res. Soc. 63, 762-770.

Brooks B.W., Barnum D.A. and Meek A.H., 1983. An Observational Study of *Corynebacterium bovis* in selected Ontario Dairy Herds. Can. J. Comp. Med. 47, 73-78.

Burvenich C., Bannerman D., Lippolis J.D., Peelman L., Nonnecke B.J., Kehrli M.E. and Paape M.J., 2007. Cumulative physiological events influence the inflammatory response of the bovine udder to Escherichia coli infections during the transition period. J. Dairy Sci. 90, E39-E54.

Burvenich C., Van Merris V., Mehrzad J. Diez-Fraile A. and Duchateau L., 2003. Severity of *E. coli* mastitis is mainly determined by cow factors. Vet. Res. 34, 521-564.

Cox D.R., 1972. Regression models and life-tables. J. R. Stat. Soc. Ser. B-Stat. Methodol. 34, 187-220.

Davidson T.J., Dohoo I.R., Donald A.W., Hariharan H. and Collins K., 1992. A cohort study of coagulase negative staphylococcal mastitis in selected dairy herds in Prince-Edward-Island. Can. J. Vet. Res. 56, 275-280.

Dego K., van Dijk J.E. and Nederbragt H., 2002. Factors involved in the early pathogenesis of bovine Staphylococcus aureus mastitis with emphasis on bacterial adhesion and invasion. A review. Vet. Quart. 24, 182-198.

de Wergifosse B., Lemaire M., Pastoret P.P. and Thiry, E., 1997. Establishment of a voluntary programme for the control of infectious bovine rhinotracheitis in the Walloon Region of Belgium. Ann. Med. Vet. 141, 185-196.

Dohoo I., Andersen S., Dingwell R. Hand K., Kelton D., Leslie K., Schukken Y. and Goddon S., 2011a. Diagnosing intramammary infections:

Comparison of multiple versus single quarter milk samples for the identification of intramammary infections in lactating dairy cows. J. Dairy Sci. 94, 5515-5522.

Dohoo I., Smith J., Andersen S., Kelton D., Goddon S., 2011b. Diagnosing intramammary infections: Evaluation of definitions based on a single milk sample. J. Dairy Sci. 94, 250-261.

Duchateau L. and Janssen P., 2008. The Frailty Model. Springer, New York.

Dufour S., Dohoo R.I., Barkema H.W., Des Côteaux L., De Vries T.J., Reyher K.K., Roy J.-P. and Scholl D.T., 2012. Manageable risk factors associated with the lactational incidence, elimination, and prevalence of Staphylococcus aureus intramammary infections in dairy cows. J. Dairy Sci. 95, 1283-1300.

Engels M. and Ackermann M., 1996. Pathogenesis of ruminant herpesvirus infections. Vet. Microbiol. 53, 3-15.

Ferguson J.D., Azzaro G., Gambina M. and Licitra G., 2007. Prevalence of mastitis pathogens in Ragusa, Sicily, from 2000 to 2006. J. Dairy Sci. 90, 5798-5813.

Goethals K., Janssen P. and Duchateau L., 2008. Frailty models and copulas: similarities and differences. J. Appl. Stat. 35, 1071-1079.

Harmon R.J., 1994. Physiology of Mastitis and Factors Affecting Somatic Cell Counts. J. Dairy Sci. 77, 2103-2112.

Hogan J. and Smith K.L., 2003. Coliform mastitis. Vet. Res. 34, 507-519.

Hogan J.S., Smith K.L., Todhunter D.A. and Schoenberger P.S., 1988. Rate of Environmental Mastitis in Quarters Infected with *Corynebacterium bovis* and Staphylococcus Species. J. Dairy Sci. 71, 2520-2525.

Heesterbeek J.A.P., 2002. A brief history of R-0 and a recipe for its calculation. Acta Biotheor. 50, 189-204.

Hsu L., Gorfine M. and Malone K., 2007. On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is mis-specified. Stat. Med. 26, 4657-4678.

Im J-K., Apley D.W., Qi C. and Shan X., 2012. A time-dependent proportional hazards survival model for credit risk analysis. J. Oper. Res. Soc. 63, 306-321.

Kaashoek M.J., van Engelenburg F.A.C., Moerman A., Gielkens A.L.J., Rijsewijk F.A.M. and van Oirschot, JT, 1996. Virulence and immunogenicity in calves of thymidine kinase- and glycoprotein E-negative bovine herpesvirus 1 mutants. Vet. Microbiol. 48, 143-153.

Kaashoek M.J. and Van Oirschot J.T., 1996. Early immunity induced by a live gE-negative bovine herpesvirus 1 marker vaccine. Vet. Microbiol. 53, 191-197.

Kaplan E.L. and Meier P., 1958. Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. 53, 457-481.

Keefe G.P., 1997. Streptococcus agalactiae mastitis: a review. Can. Vet. J. 38, 429-437.

Kerkhofs P., Renjifo X., Toussaint J.-F., Letellier C., Vanopdenbosch E. and Wellemans G., 2003. Enhancement of the immune response and virological protection of calves against bovine herpesvirus 1 with an inactivated gE-deleted vaccine. Vet. Rec. 152, 681-686.

Klein J.P. and Moeschberger M.L., 1997. Survival analysis techniques for censored and truncated data. Springer-Verlag, New York.

Kramps J.A., Perrin B., Edwards S. and van Oirschot J.T., 1996. A European inter-laboratory trial to evaluate the reliability of serological diagnosis of bovine herpesvirus 1 infections. Vet. Microbiol. 53, 153-161.

Laevens H., Deluyker H., Schukken Y. H., De Meulemeester L., Vandermeersch R., De Muelenaere E. and De Kruif A., 1997. Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. J. Dairy Sci. 80, 3219-3226.

Lam T.J.G.M., Schukken Y.H., Grommers F.J., Smit J.A.H. and Brand A., 1993. Within-herd and between-herd variation in diagnosis of clinical mastitis in cattle. J. Am. Vet. Med. Assoc. 202, 938-942.

Lam T.J.G.M., Schukken Y.H., van Vliet J.H., Grommers F.J., Tielen M.J.M. and Brand A., 1997. Effect of natural infection with minor pathogens on susceptibility to natural infection with major pathogens in the bovine mammary gland. Am. J. Vet. Res. 58, 17-22.

Lemaire M., Weynants V., Godfroid J., Schynts F., Meyer G., Letesson J.J. and Thiry E., 2000. Effects of bovine herpesvirus type 1 infection in calves with maternal antibodies on immune response and virus latency. J. Clin. Microbiol. 38, 1885-1894.

Luoma M. and Laitinen E.K., 1991. Survival analysis as a tool for company failure prediction. Omega-Int. J. Manage. Sci. 19, 673-678.

Makoschey B., Zehle H.-H., Bussacchini M., Valla G., Plfi V. and Fdi J., 2007. Efficacy of a live bovine herpesvirus type 1 marker vaccine under field conditions in three countries. Vet. Rec. 161, 295-298.

Mars M.H., De Jong M.C.M., Franken P. and Van Oirschot J.T., 2001. Efficacy of a live glycoprotein E-negative bovine herpesvirus 1 vaccine in cattle in the field. Vaccine 19, 1924-1930.

Mars M.H., De Jong M.C.M., van Maanen C., Hage J.J. and Van Oirschot J.T., 2000. Airborne transmission of bovine herpesvirus 1 infections in calves under field conditions. Vet. Microbiol. 76, 1-13.

Marshall A.W. and Olkin I., 2007. Life Distributions, Structure of Nonparametric, Semiparametric, and Parametric Families. Springer, New York.

Matthews K.R., Harmon R.J. and Langlois B.E., 1991. Effect of naturally occuring coagulase-negative Staphylococci infections on new infections by mastitis pathogens in the bovine. J. Dairy Sci. 74, 1855-1859.

Matthews K.R., Harmon R.J. and Smith, 1990. Protective effect of Staphylococcus Chromogenes infection against staphylococcus Areus infection in the lactating bovine mammary gland. J. Dairy Sci. 73, 3457-3462.

Meyer G., Hanon E., Georlette D., Pastoret P.P and Thiry E., 1998. Bovine herpesvirus type 1 glycoprotein H is essential for penetration and propagation in cell culture. J. Gen. Virol. 79, 1983-1987.

Moret-Stalder S., Fournier C., Miserez R, Albini S., Doherr M.G., Reist M., Schaeren W., Kirchhofer M., Graber H.U., Steiner A. and Kaufmann T., 2009. Prevalence study of *Staphylococcus aureus* in quarter milk samples of dairy cows in the Canton of Bern, Switzerland. Prev. Vet. Med. 88, 72-76.

Muylkens B., Thiry J., Kirten P., Schynts F. and Thiry E., 2007. Bovine herpesvirus 1 infection and infectious bovine rhinotracheitis. Vet. Res. 38, 181-209.

Muylkens B., Meurens F., Schynts F., Farnir F., Pourchet A., Bardiau M., Gogev S., Thiry J., Cuisenaire A., Vanderplasschen A. and Thiry E., 2006. Intraspecific bovine herpesvirus 1 recombinants carrying glycoprotein E deletion as a vaccine marker are virulent in cattle. J. Gen. Virol. Assoc. 87, 2149-2154.

Nardelli S., Farina G., Lucchini R., Valorz C., Moresco A., Dal Zotto R. and Constanzi C., 2008. Dynamics of infection and immunity in a dairy cattle population undergoing an eradication programme for Infectious Bovine Rhinotracheitis (IBR). Prev. Vet. Med. 85, 68-80.

Neave F.K., Dodd F.H., Kingwill R.G. and Westgarth D.R., 1969. Control of Mastitis in the Dairy Herd by Hygiene and Management. J. Dairy Sci. 52, 696-707.

Nickerson S.C. and Boddie R.L., 1994. Effect of naturally occuring coagulase-negative staphylococcal infections on experimental challange with major mastitis pathogens. J. Dairy Sci. 77, 2526-2536.

Pankey J.W., Nickerson S.C., Boddie R.L. and Hogan J.S., 1985. Effects of *Corynebacterium bovis* infection on susceptibility to major mastitis pathogens. J. Dairy Sci. 68, 2684-2693.

Parker K.I., Compton C.W.R., Anniss F.M., Heuer C. and McDougall S., 2008. Quarter-level analysis of subclinical and clinical mastitis in primiparous heifers following the use of a teat sealant or an injectable antibiotic, or both, precalving. J. Dairy Sci. 91, 169-181.

Pastoret P.P., Aguilar-Setién A., Burtonboy G., Mager J., Jetteur P. and Schoenaers F., 1979. The effect of repeated treatment with dexamethasone on the re-excretion pattern of infectious bovine rhinotracheitis virus and humoral immune response. Vet. Microbiol. 4, 149-155.

Piepers S., Opsomer G., Barkema H.W., de Kruif A. and De Vliegher S., 2010. Heifers infected with coagulase-negative staphylococci in early lactation have fewer cases of clinical mastitis and higher milk production in their first lactation than noninfected heifers. J. Dairy Sci. 93, 2014-2024.

Piessens V., De Vliegher S., Verbist B., Braem G., Van Nuffel A., De Vuyst L., Heyndrickx M. and Van Coillie E., 2012. Intra-species diversity and epidemiology varies among coagulase-negative Staphylococcus species causing bovine intramammary infections. Vet. Microbiol. 155, 62-71.

Pyörälä S., 2003. Indicators of inflammation in the diagnosis of mastitis. Vet. Res. 34, 565-578.

Pyörälä S. and Taponen S., 2009. Coagulase-negative staphylococci-Emerging mastitis pathogens. Vet. Microbiol. 134, 3-8.

Schukken Y., Chuff M.. Moroni P., Gurjar., Santisteban C., Welcome F. and Zadoks R., 2012. The "other" Gram-negative bacteria in mastitis Klebsiella, Serratia, and more. Vet. Clin. N. Am.-Food Anim. Pract. 28, 239-256.

Schukken Y.H., Grommers F.J., Van De Geer D., Erb H.N. and Brand, A., 1991. Risk Factors for Clinical Mastitis in Herds with a Low Bulk Milk Somatic Cell Count. 2. Risk Factors for *Escherichia coli* and *Staphylococcus aureus*. J. Dairy Sci. 74, 826-832.

Schukken Y.H., Günther J., Fitzpatrick J., Fontaine M.C., Goetze L., Holst O., Leigh J., Petzl W., Schuberth H.J., Sipka A., Smith D.G.E., Quesnell R., Watts J., Yancey R., Zerbe H., Gurjar A., Zadoks R.N. and Seyfert H.M., 2011. Host-response patterns of intramammary infections in dairy cows. Vet. Immunol. Immunopathol. 144, 270-289.

Schukken Y.H., Leslie K.E., Barnum D.A., Mallard B.A., Lumsden J.H., Dick P.C., Vessie G.H. and Kehrli M.E., 1999. Experimental *Staphylococcus aureus* Intramammary challenge in late lactation dairy cows: quarter and cow effects determining the probability of infection. J. Dairy Sci. 82, 2393-2401.

Schwarz D., Diesterbeck U.S., Failing K., König S., Brügemann K., Zschock M., Wolter W. and Czerny C.P. 2010. Somatic cell counts and bacteriological status in quarter foremilk samples of cows in Hesse, Germany-A longitudinal study. J. Dairy Sci. 93, 5716-5728.

Seegers H., Fourichon C. and Beaudeau F., 2003. Production effects related to mastitis and mastitis economics in diary cattle herds. Vet. Res. 34, 475-491.

Sommerhäuser J., Klopper B., Wolter W., Zschöck M., Sobiraj A. and Failing K, 2003. The epidemiology of *Staphylococcus aureus* infections from subclinical mastitis in dairy cows during a control programme. Vet. Microbiol. 96, 91-102.

Spurzem J.R., Raz M., Ito H., Kelling C.L., Stine L.C., Romberger D.J. and Rennard S.I., 1995. Bovine herpesvirus-1 infection reduces bronchial

epithelial cell migration to extracellular matrix proteins. Am. J. Physiol.-Lung Cell. Mol. Physiol. 268, L214-L220.

Stedman M.R., Lew R.A., Losina E., Gagnon D.R., Solomon D.H. and Brookhart M.A. 2011. A comparison of statistical approaches for physician-randomized trials with survival outcomes. Contemp. Clin. Trials 33, 104-115.

Supré K., Haesebrouck F., Zadoks R.N., Vaneechoutte M., Piepers S. and De Vliegher S., 2011. Some coagulase-negative Staphylococcus species affect udder health more than others. J. Dairy Sci. 94, 2329-2340.

Temple R. and Ellenberg S.S., 2000. Placebo-controlled trials and active-control trials in the evaluation of new treatments - Part 1: Ethical and scientific issues. Ann. Intern. Med. 133, 455-463.

Thiry E., Saliki J., Bublot M. and Postoret P.P., 1987. Reactivation of infectious bovine rhinotracheitis virus by transport. Comp. Immunol. Microbiol. Infect. Dis. 10, 59-63.

Tikoo S.K., Campos M. and Babiuk L.A., 1995. Bovine herpesvirus-1 (BHV-1) - biology, pahtogenisis, and Control. Academic Press inc, San Diego.

Toussaint J.F., Dubois A., Dispas M., Paquet D., Letellier C. and Kerkhofs P., 2007. Delivery of DNA vaccines by agarose hydrogel implants facilitates genetic immunization in cattle. Vaccine 25, 1167-1174.

Vaupel J.W., Manton K.G. and Stallard E., 1979. Impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16, 439-454.

Van Engelenburg F.A., Van Schie F.W., Rijsewijk F.A. and Van Oirschot J.T., 1995. Excretion of Bovine Herpesvirus 1 in semen is detected much longer by PCR than by virus isolation. J. Clin. Microbiol. 33, 308-312.

Vangroenweghe F., Dosogne H., Mehrzad J. and Burvenich C., 2001. Effect of milk sampling techniques on milk composition, bacterial contamination, viability and functions of resident cells in milky. Vet. Res. 32, 565-579.

Van Malderen G., Van Opdenbosch E. and Wellemans G., 1987. Bovine Herpes Virus 1 and 4: a sero-epidemiological survey of the Belgian cattle population. Vlaams Diergen. Tijds. 56, 364-371.

Van Oirschot J.T., Kaashoek M.J., Maris-Veldhuis M.A., Weerdmeester K. and Rijsewijk F.A.M., 1997. An enzyme-linked immunosorbent assay to detect antibodies against glycoprotein gE of bovine herpesvirus 1 allows differentiation between infected and vaccinated cattle. J. Virol. Methods 67, 23-34.

Van Schaik G., Dijkhuizen A.A., Huirne R.B.M., Schukken Y.H., Nielen M. and Hage H.J., 1998. Risk factors for existence of BHV-1 antibodies on nonvaccinating Dutch dairy farms. Prev. Vet. Med. 34, 125-136.

Watts J.L., 1988. Etiological agents of bovine mastitis. Vet. Microbiol. 16, 41-66.

Wellenberg G.J., van der Poel W.H.M. and Van Oirschot J.T., 2002. Viral infections and bovine mastitis: a review. Vet. Microbiol. 88, 27-45.

Wentinck G.H., Van Oirschot J.T. and Verhoeff J., 1993. Risk of infection with bovine herpesvirus 1 (BHV1): a review. Vet. Q. 15, 30-33.

Widodo A. and Yang B.S., 2011. Machine health prognostics using survival probability and support vector machine. Expert Syst. Appl. 38, 8430-8437.

Zadoks R.N., Allore H.G., Barkema H.W., Sampimon O.C., Wellenberg G.J., Gröhn Y.T. and Schukken Y.H., 2001b. Cow- and quarter-level risk factors for *Streptococcus uberis* and *Staphylococcus aureus* mastitis. J. Dairy Sci. 84, 2649-2663.

Zadoks R.N., Allore H.G., Hagenaars T.J., Barkema H.W. and Schukken Y.H., 2002. A mathematical model of *Staphylococcus aureus* control in dairy herds. Epidemiol. Infect. 129, 397-416.

Zadoks R.N., Gillespie B.E., Barkema H.W., Sampimon O.C., Oliver S.P. and Schukken Y.H., 2003. Clinical, epidemiological and molecular characteristics of *Streptococcus uberis* infections in dairy herds. Epidemiol. Infect. 130, 335-349.

Zadoks R.N., Middleton J.R., McDougall S., Katholm J. and Schukken Y.H., 2011. Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. J. Mammary Gland Biol. Neoplasia. 16, 357-372.

Zhang M.M., Fu S.L., Deng M.L., Xie Q., Xu H.Y., Liu Z.F., Hu C.M., Chen H.C. and Guo A.Z., 2011. Attenuation of bovine herpesvirus type

1 by deletion of its glycoprotein G and tk genes and protection against virulent viral challenge. Vaccine 29, 8943-8950.

Zhao X. and Lacasse P., 2008. Mammary tissue damage during bovine mastitis: causes and control. J. Anim. Sci. 86(Suppl. 1), 57-65.

Zhao X. and Xi J., 2011. The vaccines for Bovine Herpesvirus Type 1: A review. African Journal of Biotechnology 50, 10072-10075.

Chapter 2

Long-term effect of hyperimmunization on the transmission of infectious bovine rhinotracheitis virus in cattle herds

Based on:
Ampe B., Duchateau L., Speybroeck N., Berkvens D., Dupont A., Kerkhofs P., Thiry E. and Dispas M. (2012), "Assessment of the long-term effect of vaccination on transmission of infectious bovine rhinotracheitis virus in cattle herds hyperimmunized with glycoprotein E-deleted marker vaccine," *American Journal of Veterinary Research*, in press.

Abstract

The objective of the study is to assess the long-term effect and the effects of risk factors on the efficacy of hyperimmunization protocols against infectious bovine rhinotracheitis (IBR) during a longitudinal field study on 72 dairy and dairy-beef mixed farms with approximately 7,700 cows.

The farms were assigned to 3 treatment groups (hyperimmunization group ((HIG) 1 and 2, which were hyperimmunized with glycoprotein E (gE)-deleted marker vaccines, and a nonintervention group (NIG)). Cattle in HIG 1 were initially vaccinated with an attenuated vaccine, whereas cattle in HIG 2 were initially vaccinated with an inactivated-virus vaccine. Cattle in both HIGs received booster inoculations with inactivated-virus vaccines at approximately 6-month intervals. The hazard for gE seroconversion was compared among experimental groups via a shared frailty model with a piecewise constant baseline hazard to correct for seasonal and secular effects.

The hazard for gE seroconversion was significantly decreased over time for the 2 HIGs, compared with results for the NIG. Seasonal changes in the hazard of gE seroconversion were detected, with a higher risk in winter periods than in grazing periods. No significant difference was detected between HIG 1 and HIG 2. The only significant risk factor was the number of buildings for cattle on a farm; the higher the number of buildings, the lower the risk ratio for gE seroconversion. The mean IBR prevalence decreased over time in both HIGs but remained constant or increased in the NIG.

Hyperimmunization via repeated administration of attenuated and inactivated virus gE-deleted marker vaccines as well as inactivated-virus vaccines may provide a method for the control of IBR in cattle.

2.1 Introduction

Infectious bovine rhinotracheitis is caused by BoHV-1 and is clinically characterized by hyperemia and hypersecretion of the nasal mucosa, cough, and fever. Bovine herpesvirus 1 is also the causative agent of reproductive tract problems, such as infectious vulvovaginitis, abortions, and metritis after Cesarean section. The worldwide distribution of BoHV-1 makes it a major pathogen of cattle (Straub, 1990).

Because it is an alphaherpesvirus, BoHV-1 remains in the nervous system after infection in cattle and can be reactivated by stressful conditions (Lemaire et al., 1994). The frequency and intensity of viral shedding after

reactivation is influenced by the strain virulence and the immune status of the animal (Muylkens et al., 2007). In countries (e.g. Belgium (Boelaert et al., 2000) and the Netherlands (Van Wuijckhuise et al., 1998)) where the seroprevalence for IBR is high, the control strategy is based on repeated vaccination of all cattle in a herd. For the past 15 years, marker vaccines (Kaashoek et al., 1994; Kaashoek and Van Oirschot, 1996) have allowed for the serologic discrimination between cattle infected with BoHV-1 (seropositive against gE) and vaccinated cattle (seronegative against gE) (Van Oirschot et al., 1997). Two field studies (Mars et al., 2001; Bosch et al., 1998) have been conducted over the period of 1 year to assess efficacy for the repeated administration of gE-deleted marker vaccine to cattle in dairy herds. Both compared results for repeated vaccination with results for placebo treatments. The authors concluded that the repeated use of live-virus (Mars et al., 2001) or inactivated-virus (Bosch et al., 1998) vaccines reduce the incidence of gE seroconversion. Vaccination with a live-attenuated strain of virus resulted in a smaller reproductive ratio (Mars et al., 2001) and was used to control IBR in the Netherlands. This strategy was later rejected because of vaccine contamination (Barkema et al., 2001). More recently, hyperimmunization protocols that involve the use of live-virus vaccines have been successfully tested in 149 herds, mostly large dairy farms (Makoschey et al., 2007).

Although a hyperimmunization protocol that involves the use of sequential administration of live-virus and inactivated-virus vaccines was found to be efficient (Kerkhofs et al., 2003), it has not been tested in farm settings. Currently, putative risk factors for BoHV-1 infection have been estimated via cross-sectional observational studies (Van Wuijckhuise et al., 1998; Van Schaik et al., 1998). Only randomized studies are able to determine that a risk factor is causative for the studied disease.

For the analysis of the study reported here, a new model for survival analysis was developed. The model was designed to take into account modification of the intensity of risk factors with time, clustering of animals in a herd, seasonal differences, and the long-term change in risk over time (secular change). The objective of the study was to assess in farm settings for a 28-month period the risk factors for BoHV-1 infection in cattle herds hyperimmunized in accordance with 2 hyperimmunization protocols and compare them with those for herd without such interventions. We hypothesized that the infection rate in the HIG decreased over time compared to the NIG.

2.2 Material and Methods

2.2.1 Animals

The study was conducted on cattle from 34 dairy herds and 38 dairy-beef mixed herds that were selected from a pool of 92 volunteer herds. Inclusion criteria for the herds were that farming activity was the only source of family income and that a herd comprised less than 180 cattle.

2.2.2 Procedures

Farms were assigned to 3 groups (HIG 1, HIG 2, and NIG) by use of a randomization procedure (lottery procedure). All of the cattle in HIG 1 and 2 were hyperimmunized. This consisted of an initial 2 administrations of marker vaccines (interval of 3 to 5 weeks between administrations), which were followed by booster vaccinations at intervals of approximately 6 months. To ensure identification of the treated cattle, vaccinations were performed at the start of the study (January 1997) and then successively shortly before the start of the cattle pasture periods and as soon as possible after the start of the cattle stabling periods for a period of 28 months.

Cattle in HIG 1 initially received an intranasal administration followed by an intramuscular administration of an attenuated gE-deleted marker vaccine (Bayovac IBR-Marker vivum, at the time by Bayer AG ; Rhinobovin Marker Live, Intervet, previously Hoechst). Cattle in HIG 2 initially received 2 SC administrations of an inactivated-virus gE-deleted marker vaccine (Bayovac IBR-Marker inactivatum, then Bayer AG; Rhinobovin Marker inactivated, Intervet, previously Hoechst). Both hyperimmunized groups received boosters inoculations of the inactivated virus vaccines (Bayovac IBR-Marker inactivatum, then Bayer AG; Rhinobovin Marker inactivated, Intervet, previously Hoechst), SC, at the 6-month intervals.

For the NIG, farmers used their usual vaccination schedules. All the cattle of appropriate age for vaccination in each herd were included in the study. Only gE-deleted marker vaccines were used for all 3 groups, which allowed for serologic differentiation between infected and uninfected but vaccinated cattle (Van Oirschot et al., 1997).

2.2.3 Serologic monitoring

To limit the amount of animal handling, serologic monitoring of all cattle conducted via the same schedule as the vaccinations, which provided 6 cross-sectional blood sample collections of the 72 herds. Each blood sample was

tested with an ELISA gE antibody test kit (HerdCheck , IDEXX, USA). Inconclusive results were considered as positive **?**. Whenever possible, the serologic status of calves that had positive results because of maternal antibodies was adjusted to that of the first test value obtained after those calves were 6 months old. In all other cases, cattle that had at least 1 positive test result were definitively considered as latent carriers and positive cattle for subsequent measurements. Calves < 6 months old for which no subsequent sample was available were excluded because there was no further analysis to define their immune status. New cases of BoHV-1 infection were defined as gE seroconversion during the interval since the preceding sampling period.

2.2.4 Questionnaire for the risk factor analysis

Each owner completed a questionnaire about the herd management practices potentially linked to the risk of BoHV 1 infection; the questionnaire was completed during a face-to-face interview. The questionnaire was designed on the basis of results for studies of within-herd and between-herds risk factors (Van Schaik et al., 1998; Wentinck et al., 1993) and considered potential risk factors linked with herd management (ie, regular purchase of cattle, participation at fairs and shows, calf-dam relationship (whether calf was removed immediately after birth), use of artificial insemination or natural mating with bulls, and external sources of infection (ie, the presence of other farms within a radius of 100 m and possible contacts with other herds during the pasture period)).

2.2.5 Data analysis

Six blood collections were performed for each herd during the study. For each sample collection, seroprevalence was estimated as the ratio of the number of gE-seropositive cattle to the number of cattle from which blood samples were collected.

Data analysis was conducted to account for cattle lost to follow-up monitoring, clustering of data collected at the same farm (all cattle in a herd were managed in the same manner (random effect), season (animal density differed between winter (stabling period) and summer (grazing period)), natural turnover of cattle (birth of calves and purchase of cattle were balanced against death, culling, and sale of cattle, which resulted in a continuous entrance of cattle into and exit of cattle from the study population), and long-term change of the hazard to seroconvert over time.

A survival model is the most appropriate technique to use for analysis of

time-to-event data (such as time to seroconversion). The expected seasonal effect was corrected by assuming 2 seasons/y (ie, summer and winter), each of which had a different constant baseline risk ratio. This simple model was extended with a random effect for the farm, a frailty u_i. The shared frailty was assumed to follow a 1-parameter gamma-distribution with a mean of 1 and variance θ. This meant that an average frailty (u_i equal to 1) had no effect on the hazard for seroconversion. This resulted in a piecewise constant baseline hazard shared frailty model defined by the following equation:

$$h_{ij}(t) = \big(((\lambda_1 I(t < 153)) + (\lambda_2 I(153 < t < 335)) + ...)\, u_i \exp(\alpha + \boldsymbol{\beta}^t \boldsymbol{x}_{ijt})\big)$$

where $h_{ij}(t)$ is the hazard for seroconversion at time t for cow j on farm i with covariates x_{ijt} at time t; λ_1, λ_2, ... are the baseline risks for period 1, 2, ...; $I()$ is the indicator function, which is 1 if the condition is true and 0 if the condition is false; 153, 335, ... are the boundaries of the seasons in number of days; α is the difference in risk at the start of the study between the HIGs and the NIG; and β is the effect of an HIG over time. The baseline risk for each group was the baseline risk of the NIG multiplied by $\exp(\alpha)$. The differences at the start of the study were not expected to be significant.

Comparison of the time to seroconversion of the HIG cattle with that of the NIG was used to establish the secular change in the time to seroconversion since the start of the hyperimmunization protocol on a farm. This implied that most cattle entered the study at a time different from 0. Only the period starting at the entrance of an animal (y_{ij0}) until an event or censoring time (y_{ij}) may have an effect on the cumulative risk. Therefore, the hazard function was integrated from y_{ij0} to y_{ij} as follows:

$$H_{ij}(y_{ij}) = \int_{y_{ij0}}^{y_{ij}} h_{ij}(t)dt$$

The β values for HIG 1 and HIG 2 were an indication of the change in the risk over time. However, if hyperimmunization were to reduce the number of new infections, then a secular reduction (a negative β value for HIG 1 and HIG 2) in the risk for seroconversion would be expected.

Risk factors related to management practice that were obtained from the questionnaire were also introduced into the model as additional covariates. A forward-stepwise selection procedure was used to build the final model. Only risk factors with a value of $P < 0.1$ were considered for the stepwise procedure to build the final model. The corresponding survival function was then described by the following equation:

$$S_{ij}(t) = \big(\exp\left((\lambda_1 I(t < 153)) + (\lambda_2 I(153 < t < 335)) + ...\right) t u_i \exp(\alpha + \boldsymbol{\beta}^t \boldsymbol{x}_{ijt})\big)$$

where $S_{ij}(t)$ is the survival function at time t. By use of these equations, it was straightforward to calculate the marginal likelihood for the ith cluster as follows:

$$\begin{aligned} L_i(\lambda, \alpha, \beta, \theta) &= \int_0^\infty \prod_{j=1}^{n_i} \left(h_0(y_{ij}) u_i \exp(\alpha + \boldsymbol{\beta}^t \boldsymbol{x}_{ij})\right)^{\delta_{ij}} \\ &\quad \times \exp\left(-H_0(y_{ij}) u_i \exp(\alpha + \boldsymbol{\beta}^t \boldsymbol{x}_{ij})\right) \\ &\quad \times \frac{u^{1/\theta - 1}}{\theta^{(1/\theta)} \Gamma(1/\theta)} \exp(-\frac{u}{\theta}) du \end{aligned}$$

where $L_i(\lambda, \alpha, \beta, \theta)$ is the marginal likelihood for the ith cluster, H_0 is the cumulative baseline hazard, and δ_{ij} is the censor indicator (1 in case of an event and 0 in case of right censoring). The variance of the frailties (θ) may be interpreted as an indication of the differences among the farms (ie, the higher the variance, the greater the differences among the farms). Alternatively, after conversion to Kendall's τ (estimated as $\theta/(\theta + 2)$), the variance may be considered as a measure for the intraclass correlation (Duchateau and Janssen, 2008).

Separate models were fitted for dairy herds and dairy-beef mixed herds. Because females and males were housed in separate buildings and had different population turnover and contact structure, differences in their risk ratio for seroconversion can be assumed. Therefore, separate analyses were also performed for each sex population of the mixed herds, following the same forward-stepwise selection of risk factors.

For all final models, values of $P < 0.05$ were considered significant.

2.3 Results

2.3.1 Animals

One farmer in HIG 1 and another in HIG 2 opted to leave the study before the last winter period because they considered the schedule for collection of blood samples to be excessively restrictive. In the NIG, 12 farmers left the study during the last period because they wanted to start a hyperimmunization program or were excluded because they wanted to modify the main

type of production for their farm. Thus, at the end of the study, there were 19, 19, and 20 herds remaining in HIG 1, HIG 2, and the NIG, respectively (Table 2.1).

Table 2.1: Number of cattle (No. of herds) in each of 3 experimental groups that were tested for gE seroconversion at 6 successive blood sample collections. Cattle in HIG 1were initially vaccinated with 2 doses of an attenuated vaccine, whereas cattle in HIG 2 were initially vaccinated with 2 doses of an inactivated-virus vaccine. Then, cattle in both HIGs received booster inoculations with inactivated-virus vaccines at approximately 6-month intervals. For the NIG, farmers used their usual vaccination schedules. Only gE-deleted marker vaccines were used for all 3 groups.

Blood sample No.	HIG 1	HIG 2	NIG
1	2140 (20)	2379 (20)	3387 (32)
2	2194 (20)	2449 (20)	3337 (32)
3	2169 (20)	2383 (20)	3284 (32)
4	2152 (20)	2211 (20)	3208 (32)
5	2086 (20)	2278 (20)	3162 (32)
6	1688 (19)	1797 (19)	1718 (20)

2.3.2 Dairy herds

Although the farms were randomly allocated to the various experimental groups, the mean seroprevalence at the start of the study was higher for HIG 1 (45%) than for HIG 2 (33%) and the NIG (35%; Figure 2.1). The gE seroprevalence decreased systematically in both hyperimmunized groups during the 4 first periods, but remained almost constant in the NIG. During the last period the gE seroprevalence increased slightly for HIG 1, remained constant for HIG 2, and markedly increased in the NIG.

The estimated risks of gE seroconversion over time were calculated (Figure 2.2). At the start of the study, the baseline risks were not significantly different for HIG 1 and HIG 2, compared with that for the NIG (α values for HIG 1 and HIG 2 were not significantly different from 0; Table 2.2). The

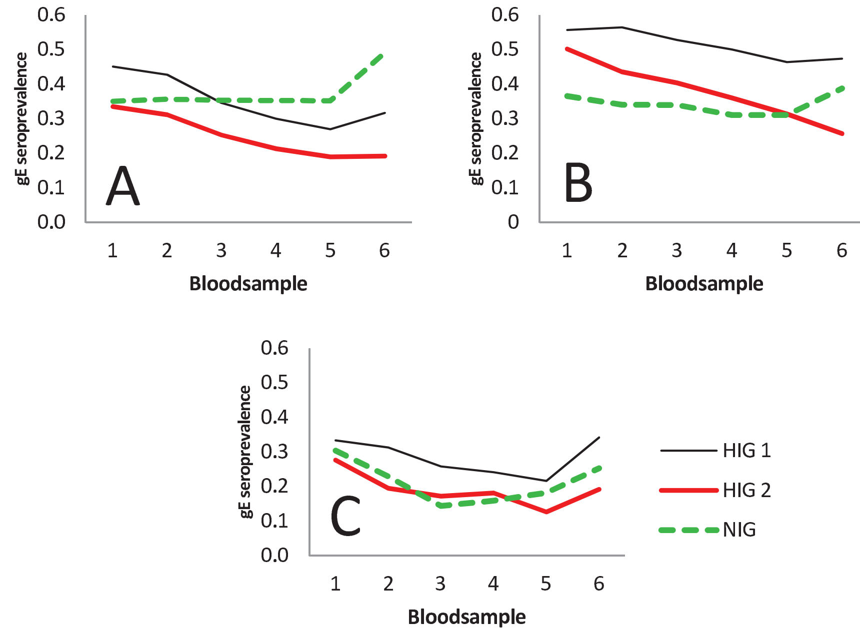

Figure 2.1: Change in the mean gE seroprevalence of cattle in 3 experimental groups (HIG 1 [thin black line], HIG 2 [thick gray line] and NIG [dashed gray line]) as determined at each of 6 blood sample collections for dairy herds (A), female cattle of dairy-beef mixed herds (B), and male cattle of dairy-beef mixed herds (C). Cattle in HIG 1 were initially vaccinated with 2 doses of an attenuated vaccine, whereas cattle in HIG 2 were initially vaccinated with 2 doses of an inactivated-virus vaccine. Then, cattle in both HIGs received booster inoculations with inactivated-virus vaccines at approximately 6-month intervals. For the NIG, farmers used their usual vaccination schedules. Only gE-deleted marker vaccines were used for all 3 groups, To minimize cattle handling, blood samples were collected at the time of vaccinations (approx 6-month intervals).

Table 2.2: The estimates for the coefficients of the piecewise constant hazard model for dairy cows. θ is the estimated variance between the herds, α_{HIG1} and α_{HIG2} are the initial differences between HIG 1 and HIG 2 compared to the NIG, and β_{HIG1} and β_{HIG2} indicate the evolution over time in years. The corresponding hazard ratio (HR) and its confidence interval is for the evolution during one year. $\beta_{nbuilding}$ is the additional effect of an extra building.

Coefficient	Estimate (se)	HR	95% CI HR	P-value
θ	1.563 (0.396)			<0.001
α_{HIG1}	-0.148 (0.566)	0.86	[0.28 ; 2.62]	0.793
α_{HIG2}	-0.568 (0.627)	0.57	[0.17 ; 1.94]	0.365
β_{HIG1}	-0.734 (0.174)	0.48	[0.34 ; 0.68]	<0.001
β_{HIG2}	-0.761 (0.197)	0.47	[0.32 ; 0.69]	<0.001
$\beta_{nbuilding}$	-0.471 (0.182)	0.62	[0.44 ; 0.89]	0.010

risk ratio for gE seroconversion was higher for the NIG than for HIG 1 or HIG 2. In addition, increased protection with time was observed for both HIG 1 and HIG 2, as indicated by a change in the risk over time in HIG 1 and HIG 2, compared with that for the NIG (β values for HIG 1 and HIG 2 were negative and significantly different from zero). Although the reduction over time was slightly greater for HIG 2 than for HIG 1, there was no significant difference in the risk ratio for gE seroconversion over time between the 2 HIGs.

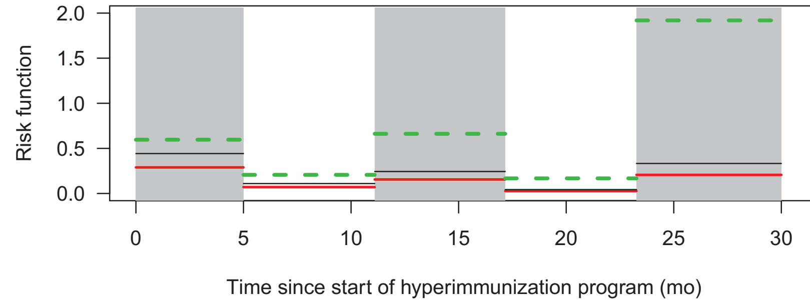

Figure 2.2: Change in the risk for gE seroconversion over time in dairy herds for HIG 1, HIG 2, and the NIG. Winter periods with stabled conditions (shaded areas) and summer conditions with grazing conditions (unshaded areas) are indicated. See Figure 1 for remainder of key.

The decreasing risk was also evident by the change in the mean seroprevalence on a herd-level basis (Figure 2.1). The seroprevalence in the HIGs primarily had a pattern of decreasing values, whereas the seroprevalence for the NIG remained almost constant.

Seasonal differences were determined by use of the risk for seroconversion for the NIG as the baseline value. The winter periods were always associated with a higher baseline risk than were the summer periods (2.2).

The estimated θ was 1.563 (Table 2.2). This corresponded to a Kendall's τ of 0.439, which indicated a high correlation within the herds.

After consideration of all available risk factors, the final model revealed that only the number of buildings for cattle on a farm had an additional significant effect on the risk ratio. The greater the number of buildings on a farm, the lower the risk for gE seroconversion (hazard ratio, 0.62/building).

2.3.3 Dairy-beef herds

Although the farms were randomly allocated among the various experimental groups, the mean seroprevalence at the start of the study for the female population was lower for the NIG (36%) than for HIG 1 (56%) and HIG 2 (46%; 2.1). At the start of the study, the baseline risk for seroconversion in females was significantly higher for the HIGs than for the NIG (α values for

Table 2.3: The estimates for the coefficients of the piecewise constant hazard model for dairy-beef herds, stratified by sex. θ is the estimated variance between the herds, α_{HIG1} and α_{HIG2} are the initial differences between P1 and P2 compared to the NIG, and β_{HIG1} and β_{HIG2} indicate the evolution over time in years. The corresponding hazard ratio (HR) and its confidence interval is for the evolution during one year.

Coefficient	Estimate (se)	HR	95% CI HR	P-value
		Cows		
θ	1.501 (0.336)			<0.001
α_{HIG1}	2.141 (0.570)	8.51	[2.78 ; 26.00]	<0.001
α_{HIG2}	1.385 (0.521)	3.99	[1.44 ; 11.09]	0.008
β_{HIG1}	-1.500 (0.189)	0.22	[0.15 ; 0.32]	<0.001
β_{HIG2}	-1.572 (0.184)	0.21	[0.14 ; 0.30]	<0.001
		Bulls		
θ	2.761 (0.946)			0.004
α_{HIG1}	0.492 (0.928)	1.64	[0.27 ; 10.08]	0.596
α_{HIG2}	-0.108 (0.783)	0.90	[0.19 ; 4.16]	0.890
β_{HIG1}	0.020 (0.414)	1.02	[0.45 ; 2.30]	0.961
β_{HIG2}	-0.639 (0.428)	0.53	[0.23 ; 1.22]	0.135

HIG 1 and HIG 2 were significantly > 0; Table 2.3). There was no significant difference in the baseline risk for gE seroconversion between HIG 1 and HIG 2. Over time, HIG 1 and HIG 2 had a large and significantly lower risk for seroconversion, compared with the risk for seroconversion for the NIG (β values for HIG 1 and HIG 2 were significantly < 0), which indicated that the protection conferred by hyperimmunization increased over time and was twice as large as the value for dairy cattle.

In the male population, there were no significant differences between the 2 HIGs and the NIG at the start of the study or over time. Also, seasonal differences were less pronounced in the male population, compared with seasonal differences in female cattle of beef-dairy mixed herds and for the dairy herds (Table 2.3).

Clustering had an important effect on the model, which indicated that herd management practices and environmental factors can have a major impact on the control of IBR. For female cattle in dairy-beef mixed herds,

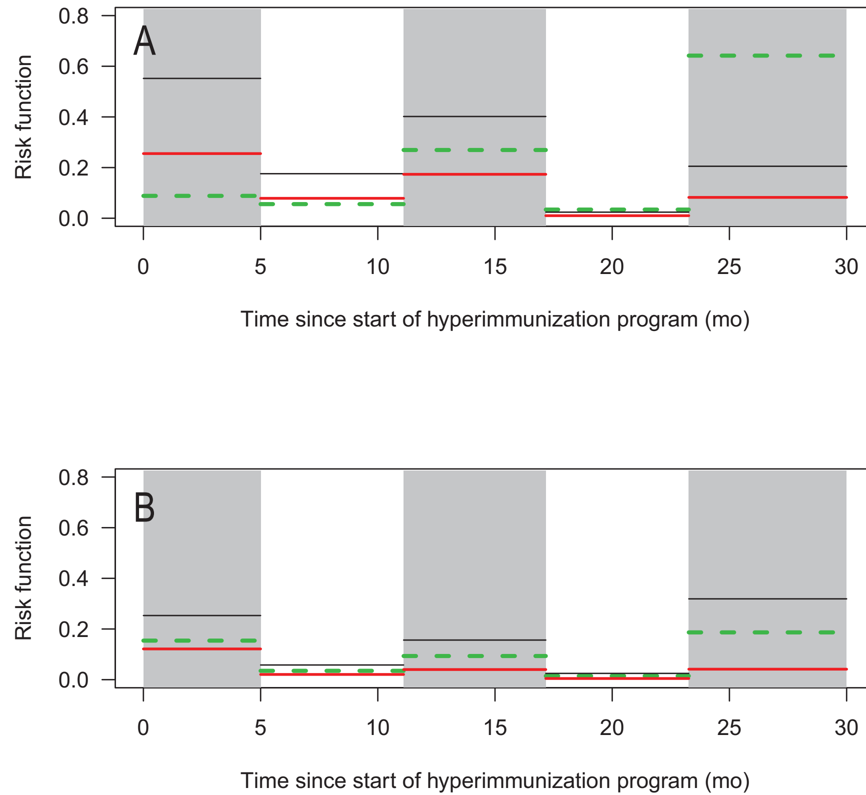

Figure 2.3: Change in the risk for gE seroconversion over time for HIG 1, HIG 2, and the NIG for females (A) and males (B) in dairy-beef mixed herds. See Figures 1 and 2 for remainder of key.

a value of 1.501 was estimated for the θ (Kendall's τ , 0.429), which was similar to the values for the dairy herds. However, the estimated θ in beef bulls was 2.761, which was much higher than that of the female cattle, and resulted in an extremely high intraclass correlation (τ = 0.580).

All risk factors were considered for the final model of female and male cattle of the diary-beef mixed herds. However, none of them had a significant effect on the hazard for gE seroconversion.

2.4 Discussion

The randomized longitudinal study reported here was based on the individual gE serologic follow up monitoring of all cattle in 3 experimental groups on dairy and dairy-beef mixed farms. Two hyperimmunization protocols were used, and both significantly reduced the risk of BoHV 1 infection in the females, when compared with the risk for the NIG. However, it was not possible to detect the same effect in the male population of the dairy-beef mixed herds. These results are consistent with those obtained in previous field studies that revealed a reduction of virus circulation in cattle herds hyperimmunized with live-virus vaccine (Mars et al., 2001; Makoschey et al., 2007) or an inactivated-virus vaccine (Bosch et al., 1998).

It is not possible to make further comparisons between results of other studies and those of the present study because conclusions of other studies were based on the basic reproduction number (Diekmann et al., 1990) calculated for a period of 18 months in studies in which only female cattle more than 1 year old were included. By comparison, the present study was based on an analysis of the time to seroconversion in all cattle on the farms measured over shorter intervals and for a much longer period (28 months). In contrast to the calculation of the basic reproduction number, the proposed piecewise constant risk model estimates the change in seroprevalence over time and takes into account clustering, different levels of risk over time, seasonal effects, and the development of herd immunity. This results in more information about the behavior of the infection, immunity on a herd-level basis, and differences between herds than does only a single summary measure such as the reproductive ratio.

Selection of the 2 hyperimmunization protocols was based on results of previous studies. The protocol for HIG 1 was based on the conclusions of an experimental trial (Kerkhofs et al., 2003) that revealed successive administration of live-virus and inactivated-virus vaccines induced the production of antibodies and a cellular immune response, which were significantly higher than those obtained with only a single administration. The protocol for HIG 2 was based on the administration of an inactivated-virus vaccine, which had already been tested in a field setting, but for a shorter period (Bosch et al., 1998). Both protocols were designed to prevent the development of major problems linked with the intensive use of live-attenuated virus vaccines. These problems include contamination of the vaccine by another virus, reactivation or re-excretion of a gE-negative vaccine strain in a field setting (Dispas et al., 2003), and potential emergence of highly pathogenic gE-negative mutant strains of BoHV-1 (Muylkens et al., 2006) as a result of

recombination between vaccine and field strains.

A NIG, rather than a placebo treatment, was used in the present study, even though it is more difficult to detect a significant difference between control and treatment groups. A placebo treatment was not acceptable because of the increased risk of virus circulation and infection, comparison with the risk for the usual management procedures. In addition, it is unethical to force farmers to stop vaccinating cattle (ie, placebo treatment) for several years if their animals are selected for the control group. Thus, an NIG is commonly used to compare new treatments with standard treatments (Djulbegovic and Clarke, 2001; Pocock, 2003).

In dairy herds, there were no differences in the initial risk, but the change over time was significantly < 0 for the HIGs. This resulted in a decrease in risk over time for the HIGs, compared with the NIG. The risk of gE seroconversion for HIG 1 was always higher than, or equivalent to, the HIG 2 (Figure 2); however, the differences were not significant. This observation is in agreement with results of a study (Bosch et al., 1998) in which animals hyperimmunized with inactivated-virus vaccines shed less virus after reactivation than did animals vaccinated with live-virus vaccine. This suggests that the most efficacious approach for the eradication of BoHV-1 is the use of inactivated-virus vaccines to reduce reactivation, as opposed to the administration of live-virus vaccines in an attempt to prevent new infections. This strategy is also supported by results of another study (Bosch, 1997).

The increased risk for seroconversion during the winter periods, which was caused by commingling of cattle of various age classes and infection status in barns, revealed the limits of hyperimmunization in stressful situations. To be effective, hyperimmunization has to be applied for a long period (as indicated by the negative β_{HIG1} and β_{HIG2}; Tables 2.2 and 2.3). Even then, virus circulation cannot be totally stopped. Therefore, additional biosecurity measures should accompany the use of hyperimmunization protocols. First, adult cows and heifers should receive booster inoculations of inactivated-virus gE-deleted IBR vaccine at least 15 days before the end of the grazing period. Second, whenever possible during the winter period, gE-seropositive cattle should be separated from cattle seronegative for gE. The study revealed that there was a significant decrease in the risk ratio for seroconversion when a farm had more buildings for cattle. In addition, as a general rule, calves should be immediately removed from their dams, and young stock should not have contact with dry (nonlactating) cows. This will allow for complete initial vaccination and booster administration before contact with a group of older animals, which are potential virus shedders. The specific accelerated removal of gE-seropositive cattle should be used to

decrease the length of the hyperimmunization schedule.

In dairy-beef mixed herds, there was a difference between the female and male populations. For the female population, the findings were in agreement with those obtained for the dairy herds, although the risk ratio for seroconversion was significantly larger in the HIGs than in the NIG at the start of the study. This is probably attributable to the higher seroprevalence, and corresponding infection pressure, at the start of the study.

The risk of gE seroconversion in the HIGs decreased over time, compared with that for the NIG, and there was no significant difference between the 2 HIGs. The decrease over time in the females of the dairy-beef mixed herds was faster than that of the dairy herds. The higher seroprevalence for the HIGs at the start of the study is a possible explanation for this.

For the male population, we did not detect a significant difference in the change of the risk for gE seroconversion over time and the seroprevalence among the 3 groups, which had almost the same pattern. This may have been associated with an extremely high turnover rate associated with the management of feedlots (grouping of stressed cattle without previous testing for IBR and possible infection before optimal protection from vaccination). A reduction in infection pressure can be expected when the number of cattle shedding virus decreases among older animals as a result of the hyperimmunization. In this category a high clustering of infection times at the farm level was observed (high θ; Table 2.3), which resulted in a higher correlation within the herds for the male population than for the female population. This is indicative of large differences among the farms and suggested that the management of each farm and environmental factors can have a major impact on the control of IBR.

2.5 Conclusion

In the present study, we suggested that hyperimmunization that involved repeated administration of attenuated and inactivated-virus gE-deleted marker vaccine (HIG 1) as well as inactivated vaccines (HIG 2) allowed for the control of IBR. The efficacy of vaccinations can be maintained by appropriate management during the winter (cold season) and thus shorten the time until eradication of IBR.

Bibliography

Barkema H., Bartels C., Van Wuijckhuise L., Hesselinck J., Holzhauer M., Weber M., Franken P., Kock P., Bruschke C. and Zimmer G., 2001. Outbreak of bovine virus diarrhea on Dutch dairy farms induced by a bovine herpesvirus 1 marker vaccine contamined with bovine viral diarrhea virus type 2. Tijdschr. Diergeneeskd. 126, 158-165.

Boelaert F., Biront P., Soumare B., Dispas M., Vanopdenbosch E., Vermeersch J.-P., Raskin A., Dufey J., Berkvens D. and Kerkhofs P., 2000. Prevalence of bovine herpesvirus-1 in the Belgian cattle population. Prev. Vet. Med. 45, 285-295.

Bosch J.C., De Jong M., Franken P., Frankena K., Hage J., Kaashoek M.J., Maris-Veldhuis M., Noordhuizen J., Van Der Poel W., Verhoef J., Weerdmeester K., Zimmer G. and Van Oirschot J.T., 1998. An inactivated gE-negative marker vaccine and an experimental gD-subunit vaccine reduce the incidence of bovine herpesvirus 1 infections in the field. Vaccine 16, 265-271.

Bosch J.C., Kaashoek M.J. and van Oirschot J.T., 1997. Inactivated bovine herpesvirus 1 marker vaccines are more efficacious in reducing virus excretion after reactivation than a live marker vaccine. Vaccine 15, 1512-1547.

de Wergifosse B., Lemaire M., Pastoret P.P. and Thiry, E., 1997. Establishment of a voluntary programme for the control of infectious bovine rhinotracheitis in the Walloon Region of Belgium. Ann. Med. Vet. 141, 185-196

Diekmann O., Heesterbeek J. and Metz J., 1990. On the definition and the computation of the basic reproduction ratio (R0) in models for infectious diseases in heterogeneous populations. J. Math. Biol. 28, 365-382.

Dispas M., Schynts F., Lemaire M., Letellier C., Vanopdenbosch E., Thiry E. and Kerkhofs P., 2003. Isolation of a glycoprotein E deleted bovine herpesvirus-1 strain in the field. Vet. Rec. 152, 209-212.

Djulbegovic B. and Clarke M., 2001. Scientific and ethical issues in equivalence trials. J. Am. Med. Assoc. 9, 1206-1208.

Duchateau L. and Janssen P., 2008. The Frailty Model. Springer, New York.

Kaashoek M.J., Moerman A., Madic J., Rijsewijk F.A., Quak J., Gielkens A.L. and Van Oirschot J.T., 1994. A conventionally attenuated glycoprotein E-negative strain of bovine herpesvirus type-1 is an efficacious and safe vaccine. Vaccine 12, 439-444.

Kaashoek M.J. and Van Oirschot J.T., 1996. Early immunity induced by a live gE-negative bovine herpesvirus 1 marker vaccine. Vet. Microbiol. 53, 191-197.

Kerkhofs P., Renjifo X., Toussaint J.-F., Letellier C., Vanopdenbosch E. and Wellemans G., 2003. Enhancement of the immune response and virological protection of calves against bovine herpesvirus 1 with an inactivated gE-deleted vaccine. Vet. Rec. 152, 681-686.

Kramps J.A., Perrin B., Edwards S. and van Oirschot J.T., 1996. A European inter-laboratory trial to evaluate the reliability of serological diagnosis of bovine herpesvirus 1 infections. Vet. Microbiol. 53, 153-161.

Lemaire M., Pastoret P.-P. and Thiry E., 1994. Le contrôle de l'infection par le virus de la rhinotrachite infectieuse bovine. Ann. Med. Vet. 138, 167-180.

Makoschey B., Zehle H.-H., Bussacchini M., Valla G., Plfi V. and Fdi J., 2007. Efficacy of a live bovine herpesvirus type 1 marker vaccine under field conditions in three countries. Vet. Rec. 161, 295-298.

Mars M.H., De Jong M.C.M., Franken P. and Van Oirschot J.T., 2001. Efficacy of a live glycoprotein E-negative bovine herpesvirus 1 vaccine in cattle in the field. Vaccine 19, 1924-1930.

Muylkens B., Meurens F., Schynts F., Farnir F., Pourchet A., Bardiau M., Gogev S., Thiry J., Cuisenaire A., Vanderplasschen A. and Thiry E., 2006. Intraspecific bovine herpesvirus 1 recombinants carrying glycoprotein E deletion as a vaccine marker are virulent in cattle. J. Gen. Virol. Assoc. 87, 2149-2154.

Muylkens B., Thiry J., Kirten P., Schynts F. and Thiry E., 2007. Bovine herpesvirus 1 infection and infectious bovine rhinotracheitis. Vet. Res. 38, 181-209.

Pocock S.J., 2003. The pros and cons of noninferiority trials. Fundam. Clin. Pharmacol. 17, 483-490.

Straub O.C., 1999. Infectious bovine rhinotracheitis virus. In: Dinter Z, Morein B, eds. Virus infections of ruminants. Amsterdam: Elsevier Science Publishers, 1990, 71-108.

Van Oirschot J.T., Kaashoek M.J., Maris-Veldhuis M.A., Weerdmeester K. and Rijsewijk F.A.M., 1997. An enzyme-linked immunosorbent assay to detect antibodies against glycoprotein gE of bovine herpesvirus 1 allows differentiation between infected and vaccinated cattle. J. Virol. Meth. 67, 23-34.

Van Schaik G., Dijkhuizen A.A., Huirne R.B.M., Schukken Y.H., Nielen M. and Hage H.J., 1998. Risk factors for existence of BHV-1 antibodies on nonvaccinating Dutch dairy farms. Prev. Vet. Med. 34, 125-136.

Van Wuijckhuise L., Bosch J.C., Franken P., Frankena K. and Elbers A.R., 1998. Epidemiological characteristics of bovine herpesvirus 1 infections determined by bulk milk testing of all Dutch dairy herds. Vet. Rec. 142, 181-184.

Wentinck G.H., Van Oirschot J.T. and Verhoeff J., 1993. Risk of infection with bovine herpesvirus 1 (BHV1): a review. Vet. Q. 15, 30-33.

Chapter 3

Investigating clustering in interval-censored udder quarter infection times in dairy cows using a gamma frailty model

Based on:
Ampe, B., Goethals, K., Laevens, H. Duchateau, L. (2012), "Investigating clustering in interval-censored udder quarter infection times in dairy cows using a gamma frailty model," *Preventive Veterinary Medicine*, 106, 251-257.

Abstract

Udder infections in dairy cows are observed at udder quarter level. Therefore, the best strategy to study infection dynamics of particular bacteria causing mastitis is to follow up and model individual udder quarter infection times. Udder quarter infection times, however, are not independent as they are clustered within a cow and herds. Another challenge in modelling infection times is that the exact infection time is unknown; it is only known that the infection has taken place in the interval between the last negative and the first positive sample. We applied a technique based on the gamma frailty model which handles the clustering and interval censoring simultaneously. Parameter estimates can be obtained analytically and their variance is obtained by the inverse of the hessian matrix. The proposed technique was applied to udder quarter infection times for *Corynebacterium bovis*, *Staphylococcus aureus* and *Streptococcus uberis*. Multiparous cows were more likely to get infected earlier in lactation with *C. bovis* or *S. uberis* than primiparous cows. The times to infection of all three bacteria were highly clustered at cow level and the results of a stratified model on a subset of herds suggested a high clustering on herd level for *C. bovis* and *S. uberis*.

3.1 Introduction

Mastitis is one of the major health disorders in dairy cattle. It has a negative effect on the milk yield and milk quality (increase of somatic cell count). The economic impact results from the control costs (i.e. extra resource use, treatment,...) and losses due to reduced revenues (Seegers et al., 2003). Mastitis is a multifactorial disease. Numerous factors (breeding practices, hygiene, quality of milking, weather, genotype, calving month, lactation stage, ...) have an effect on the susceptibility to udder infections with major, minor and facultative pathogens and their ability to induce mastitis, not every intramammary infection results in a clinical mastitis (Gasqui et al., 2000).

Intramammary infections (IMI) have most often been modelled using Poisson regression with the number of new infections as response variable (Zadoks et al., 2001b). Allore and Erb (1999) and Zadoks et al. (2002) used a SIR-model (Susceptible, Infected, Recovered) to estimate the transition parameters from susceptible/uninfected to infected and from recovered to infected. Udder quarters were treated as independent entities. None of the models above takes into account the actual time to infection with the available information reduced to a binary variable.

Gasqui et al. (2000) introduced survival analysis as a tool to model clinical mastitis data. A piecewise constant hazard model was used to model the time intervals between two consecutive cases of clinical mastitis during the same lactation.
If infections are followed up at quarter level, however, more complex techniques have to be considered due to the fact that udder quarters are clustered within cow, and cows are even clustered within herds. Furthermore, the infection times are not known exactly. It is only known that the infection occurred between the last negative and the first positive test. This is often the case in observational studies because it is not feasible to assess the infection status of each udder quarter on a daily basis. The standard statistical tools need therefore to be extended to cope with this type of missing information, called interval censoring. In this paper a statistical technique is proposed that can handle the clustering and interval censoring in the data simultaneously. This technique will be used to investigate the effect of lactation and the degree of clustering for three well-known intramammary bacteria, *Corynebacterium bovis*, *Staphylococcus aureus* and *Streptococcus uberis*.

3.2 Materials and methods

3.2.1 Animal database

Twenty-five dairy herds, located in East and West Flanders (Belgium), were followed up during a 20-months period (February 1993 to September 1994) (Laevens et al., 1997). Only lactations with the first day of the lactation between February 1993 and September 1994 were considered.
At monthly intervals, 11 times a year (herds were not visited during either July or August), quarter foremilk samples were taken to detect IMI. More details about the isolation and identification procedure are given by Laevens et al. (1997).
Different bacteria were followed up. In this article we focus on three of them. *S. aureus* and *S. uberis* are two well-known pathogens. *S. aureus* is a contagious major pathogen while *S. uberis* is an environmental major pathogen. An infection with one of those two pathogens results most likely in clinical mastitis. *C. bovis* is a contagious minor pathogen, which normally does not result in clinical mastitis (Schukken et al., 1999; Bradley, 2002).

3.2.2 Statistical methods

Individual udder quarter infection times can be modelled most efficiently by survival analysis techniques because some infection times were right censored due to the fact that no infection occurred during the lactation period or that a cow was culled before she experienced an IMI. It is also likely that infection times of the four udder quarters of the same cow are correlated. To take the clustering at the cow level into consideration, a frailty term for cow was added to the model. The modelling was based on the parametric proportional hazards shared frailty model with a baseline hazard assumed to follow a Weibull distribution. The hazard ($h_{ij}(t)$) at time t for quarter j ($j = 1, 2, 3, 4$) from cow i ($i = 1, \ldots, n$) is given by:

$$h_{ij}(t) = \lambda \gamma t^{\gamma-1} u_i \exp(\boldsymbol{\beta}^t \boldsymbol{x}_{ij})$$

with u_i the effect of cow i assumed to come from a one-parameter gamma density with mean one and variance θ, covariates $\boldsymbol{x}_{ij}^t = (x_{ij1}, x_{ij2} \ldots, x_{ijp})$ and parameters $\boldsymbol{\beta}^t = (\beta_1, \beta_2, \ldots, \beta_p)$. In the presented analysis, only parity (primiparous or multiparous) was considered as covariate. The baseline hazard is assumed to follow a Weibull distribution with scale parameter λ and shape parameter γ. The survival function corresponding to this hazard function is given by

$$S_{ij}(t) = \exp\left(-\lambda t^{\gamma} u_i \exp(\boldsymbol{\beta}^t \boldsymbol{x}_{ij})\right)$$

with u_i the effect of cow i assumed to come from a one-parameter gamma density with mean one and variance θ:

$$f_U(u) = \frac{u^{(1/\theta)-1} \exp(-u/\theta)}{\theta^{(1/\theta)} \Gamma(1/\theta)}$$

For interval-censored data, the true infection time is only known to be between the last observed time without infection (lower bound) and the first time with infection (upper bound). Ignoring interval censoring results in bias (Radke, 2003). The probability that the true time of infection is situated between the lower (L_{ij}) and upper (U_{ij}) bound of the interval is given by the difference between the survival probability at the lower bound and the survival probability at the upper bound (Collet, 2003; Duchateau and Janssen, 2008). For right censored data, the last observed time is given by R_{ij}. Therefore, the likelihood for data with both right and interval censoring is given by

$$L(\boldsymbol{\beta}, \lambda, \gamma) = \prod_{i=1}^{n} \prod_{j=1}^{4} \left(S_{ij}(L_{ij}) - S_{ij}(U_{ij})\right)^{\delta_{ij}} \left(S_{ij}(R_{ij})\right)^{(1-\delta_{ij})}$$

where δ_{ij} is the censoring indicator (0: censored, 1: event, interval censored).

The conditional likelihood (conditional on the frailties) is given by

$$\begin{aligned} L(\boldsymbol{\beta}, \lambda, \gamma, \mathbf{u}) &= \prod_{i=1}^{n} \prod_{j=1}^{4} \Bigg[\Big[\exp\left(-u_i \lambda L_{ij}^{\gamma} \exp(\boldsymbol{\beta}^t x_{ij})\right) \\ &\quad - \exp\left(-u_i \lambda U_{ij}^{\gamma} \exp(\boldsymbol{\beta}^t x_{ij})\right)\Big]^{\delta_{ij}} \\ &\quad \Big[\exp\left(-u_i \lambda R_{ij}^{\gamma} \exp(\boldsymbol{\beta}^t x_{ij})\right)\Big]^{1-\delta_{ij}} \Bigg] \end{aligned}$$

The conditional likelihood contains the unobserved frailty terms $\boldsymbol{u} = (u_1, \ldots, u_n)$. Goethals et al. (2009) demonstrated that the frailties can be integrated out analytically from the likelihood of interval-censored frailty models if the gamma density is assumed for the frailties. The resulting marginal likelihood does no longer contain the frailties, but only the variance of the frailties θ. This is an extension of the results obtained for right censored survival (Klein and Moeschberger, 1997; Duchateau et al., 2002) to interval-censored data. The model without interval censoring is further referred to as the naive model, the model extended to interval censoring is the proposed model.

The marginal likelihood is given by:

$$\begin{aligned} L(\boldsymbol{\beta}, \lambda, \gamma) &= \prod_{i=1}^{n} \int_{0}^{\infty} \prod_{j=1}^{4} \Bigg[\Big[\exp\left(-u \lambda L_{ij}^{\gamma} \exp(\boldsymbol{\beta}^t x_{ij})\right) \\ &\quad - \exp\left(-u \lambda U_{ij}^{\gamma} \exp(\boldsymbol{\beta}^t x_{ij})\right)\Big]^{\delta_{ij}} \\ &\quad \Big[\exp\left(-u \lambda R_{ij}^{\gamma} \exp(\boldsymbol{\beta}^t x_{ij})\right)\Big]^{1-\delta_{ij}} \Bigg] \frac{u^{(1/\theta)-1} \exp(-u/\theta)}{\theta^{(1/\theta)} \Gamma(1/\theta)} \mathrm{d}u \end{aligned}$$

and can be maximized by general purpose maximizers such as the nlm function in R, which further supplies the hessian matrix. The inverse of the hessian matrix is an estimate of the variance-covariance matrix of the estimated parameters $\boldsymbol{\zeta} = (\theta, \lambda, \gamma, \boldsymbol{\beta})$.

In most cases β is transformed to the conditional hazard ratio ($\mathrm{HR} = \exp(\beta)$), conditional on the same frailty.

The variance of the frailties (θ) is of inherent interest. It is however not straightforward to interpret the parameter θ. It can be understood in two different ways.

First the presence of frailties (with their variance) creates correlation between the infection times of the four quarters within a cow udder. For the gamma frailty model, the non-parametric intraclass correlation coefficient, known as Kendall's τ, is given by $\theta/(2+\theta)$ (Glidden and Vittinghoff, 2004).

Kendall's τ is a global measure of dependence defined by $\tau = \mathrm{E}[\mathrm{sign}\{(T_{ij} - T_{lk})(T_{ij'} - T_{lk'})\}]$, where $(T_{ij}, T_{ij'})$ and $(T_{lk}, T_{lk'})$ are two randomly chosen pairs of infection times from two randomly chosen cows i and l and $\mathrm{sign}(x) = -1, 0, 1$ for $x < 0, x = 0, x > 0$ respectively. A pair is concordant if $(T_{ij} - T_{lk})(T_{ij'} - T_{lk'}) > 0$ and otherwise discordant (Oakes, 1989). This gives an idea about the correlation between the four quarters of the same cow. If one quarter is infected before a quarter of another cow, the probability that another quarter of the same cow is also infected, before another quarter of the other cow, increases with increasing values of Kendall's τ.

Second, the presence of the frailties causes the hazard function to differ from cow to cow. The frailty hence operates at the level of the hazard, which makes interpretation difficult. It can, however, be translated in terms of variability of median infection time (t_m) between cows (Duchateau and Janssen, 2005). The density function (f_{m_i}) for the median time to infection for cow i with covariates $\boldsymbol{x}_i$ is given by the following expression

$$f_{m_i}(t_m) = \gamma \left(\frac{\log 2}{\theta \lambda \exp(\boldsymbol{\beta}^t \boldsymbol{x}_i)} \right)^{1/\theta} \frac{1}{\Gamma(1/\theta)} \left(\frac{1}{t_m} \right)^{1+\frac{\gamma}{\theta}} \exp\left(- \frac{\log 2}{\theta t_m^{\gamma} \lambda \exp(\boldsymbol{\beta}^t \boldsymbol{x}_i)} \right) \tag{3.1}$$

If only a small proportion of the cows has experienced an event, the meaning of the median time to infection is not very useful. The density function (3.1) can therefore be adapted to a certain quantile instead of the median. In our setting the 5% quantile is used and can be interpreted as the time at which 5% of the quarters have experienced an intramammary infection. The density function (3.1) can be adapted to a more general form where Q_i is the considered quantile for cow i

$$f_{Q_i}(t_Q) = \gamma \left(\frac{\log(\frac{1}{1-Q})}{\theta \lambda \exp(\boldsymbol{\beta}^t \boldsymbol{x}_i)} \right)^{1/\theta} \frac{1}{\Gamma(1/\theta)} \left(\frac{1}{t_Q} \right)^{1+\frac{\gamma}{\theta}} \exp\left(- \frac{\log(\frac{1}{1-Q})}{\theta t_Q^{\gamma} \lambda \exp(\boldsymbol{\beta}^t \boldsymbol{x}_i)} \right)$$

The variability of the frailties in the proposed model is due to differences between cows at population level. But also differences between the herds can

have their influence. To investigate the variability between cows within each herd a stratified analysis was performed. In some herds the analysis failed due to a lack of events (less than 2% events). Those herds were excluded for the stratified analysis. The stratified analysis results in frailties on cow level within the same herd. The stratified analysis was performed in two steps, in the first step the fixed effect was considered to be the same in each herd, the variance of the frailties and the Weibull distribution parameters (λ and γ) for the baselinehazard were allowed to differ at herd level. In the next step this model was reduced to a model where the fixed effect and the variance of the frailties on cow level were assumed to be the same within each herd, only the baselinehazard was allowed to differ at herd level. The two models were compared using the likelihood ratio test. The last model was also compared to the unstratified model fitted on the same subset of herds.

3.3 Results

During the study period, 1132 cows were followed in 25 different herds. This resulted in quarter milk samples of 4526 quarters. *C. bovis* was most frequently isolated in the quarter milk samples (1873 times or 39% of the quarters, only the first positive sample was counted) followed by *S. aureus* (228 - 5%) and *S. uberis* (177 - 4%). The evolution of the hazard function of the proposed model during the lactation is depicted in Figure 3.1 for the three bacteria for cows with frailty equal to 1 (which is an average cow). The hazard increased with time for all three bacteria, but mostly so for *C. bovis*, with γ equal to 1.975 (Table 3.1). For *S. uberis* and *S. aureus*, γ was close to one (constant hazard over time) and did not differ significantly from one.

The estimated fixed effect of multiparous versus primiparous for time to *C. bovis* infection was 0.867 (s.e.: 0.14) which corresponds to a hazard ratio of 2.38 (95% CI: [1.81;3.13]), meaning that a multiparous cow was 2.38 times more likely to be infected than primiparous cows at any time. The hazard ratio for *S. uberis* was slightly smaller and equal to 2.32 (95%CI [1.60;3.34]). For *S. aureus*, there was only a small difference between multiparous and primiparous cows with the hazard ratio equal to 1.30 (95%CI [0.86;1.96]), which did not differ significantly from one.

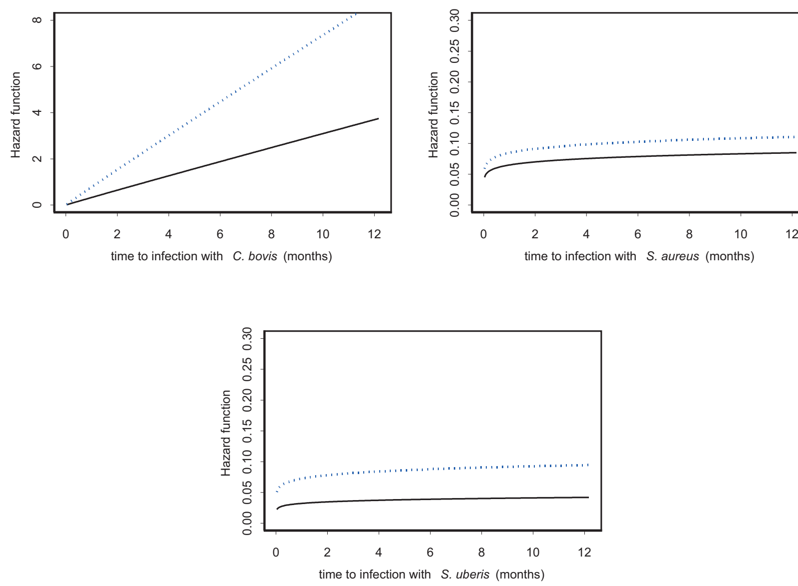

Figure 3.1: Hazard plot of first infection in months for primiparous cows (solid line) and multiparous cows (dashed line)

Table 3.1: The estimates (and standard errors) of the parameters in the proposed (interval censoring) and the naive model (midpoint imputation). The (*) indicates that the variance of the frailties (θ), the scale parameter (λ) or the fixed effect for parity (β) is significantly different from 0 or from 1 in case of the shape parameter (γ).

	parameter	Proposed model	Naive model
C. bovis	θ (se)	3.787* (0.220)	3.651* (0.214)
	λ (se)	0.0140* (0.0017)	0.0130* (0.013)
	γ (se)	1.975* (0.042)	2.060* (0.041)
	β (se)	0.867* (0.142)	0.887* (0.141)
S. aureus	θ (se)	5.578* (0.907)	5.639* (0.919)
	λ (se)	0.0049* (0.0009)	0.0043* (0.0007)
	γ (se)	1.107 (0.069)	1.207* (0.073)
	β (se)	0.265 (0.208)	0.275 (0.211)
S. uberis	θ (se)	5.852* (1.084)	5.757* (1.072)
	λ (se)	0.0024* (0.0007)	0.0022* (0.0005)
	γ (se)	1.106 (0.078)	1.199* (0.081)
	β (se)	0.811* (0.241)	0.827* (0.239)

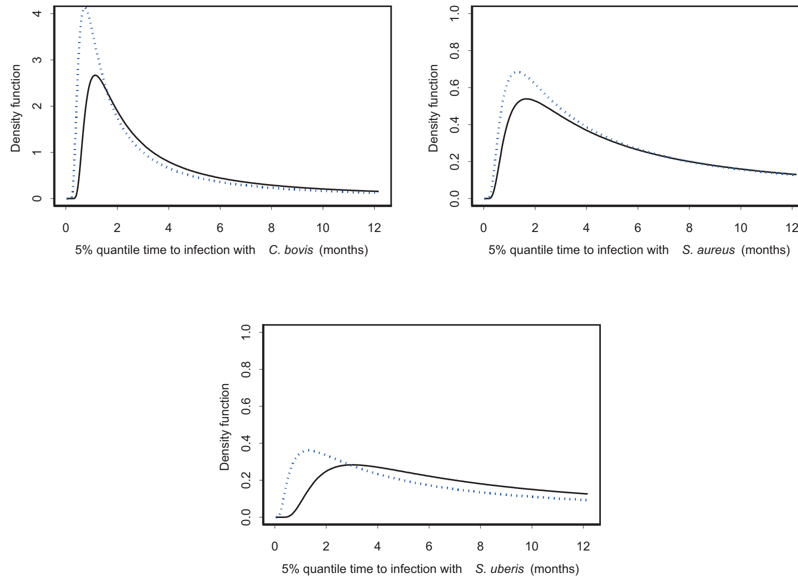

Figure 3.2: 5% quantile time to first infection in months (solid line: primiparous cows, dashed line: multiparous cows)

The median time to *C. bovis* infection for a quarter of a primiparous (multiparous) cow with a frailty equal to 1 was 7.2 (3.9) months. The median time to *S. aureus* and *S. uberis* infection was not available as less than 50 % of the cows was infected with these bacteria during the lactation period. The 5% quantile of *C. bovis* was 1.9 (1.2) months. The 5% quantile of *S. aureus* was 8.3 (6.6) months and for *S. uberis* 15.6 (7.5) months. This means that only a small part of the quarters experienced an intramammary infection within one lactation. Multiparous cows had a much lower 5% quantile time than the primiparous cows.

The variance of the frailties was in all three cases very high, for *C. bovis* the variance was 3.78 (s.e.: 0.22), for *S. aureus* 5.578 (s.e.: 0.91) and for *S. uberis* 5.852 (s.e.: 1.08). These variances corresponded to a Kendall's τ of 0.65 for *C. bovis*, 0.74 for *S. aureus* and 0.75 for *S. uberis*. The high variance of the frailties results in a widely spread distribution of the 5% quantile time to infection of individual cows. The density functions are given in Figure 3.2. The effect of the high variance of the frailties was less pronounced for

the multiparous cows due to the higher hazard rate.
To illustrate the bias when ignoring interval censoring, the results of the frailty model without interval censoring (naive model, midpoints of the intervals were considered as true event times) are compared with the proposed model in Table 3.1.
The estimates for θ and standard errors are similar for both models. The fixed effect is slightly overestimated in the naive model.
The Weibull scale-parameter (λ) is underestimated and the shape-parameter (γ) is overestimated in the naive model. This results in an underestimation of the baseline hazard and a steeper increase during lactation. The results of the proposed model still indicates an increasing hazard (Figure 3.1) for intramammary infection. Gasqui et al. (2000) also found an increasing hazard for mastitis during one lactation.

In the stratified analysis, herds with only a few or no events were excluded (for *C. bovis* 5, for *S. aureus* 12 and for *S. uberis* 9 herds were excluded), which resulted in an overestimation of the baseline hazard. It has to be stressed that the conclusions of the stratified model are limited to the herds in the subset and do not include all herds. In the first step a stratified model was fitted where θ, λ and γ could differ in each herd. In the next step, the variance of the frailties was considered to be the same in each herd. for *S. aureus* (P=0.87) and for *S. uberis* (P=0.22) the likelihood ratio test did not indicate a significant difference between the full en reduced model which indicates that the variance on cow level could be assumed to be the same in all herds. In the case of *C. bovis*, a reduction to one variance of the frailties was tested and indicated a significant difference between the herds (P<0.01). The next results are based on the model where the fixed effect en variance of the frailties was assumed to be the same for all herds.

The frailties of the stratified model included only cow effects within the same herd. The estimated variance of the frailties of the stratified model decreased compared to the unstratified model (fitted on the same subset of herds) (*C. bovis*: unstratified: 3.787 - stratified: 1.29, *S. aureus*: 5.578 - 2.699 and *S. uberis*: 5.852 - 3.989) To give an impression of the differences between herds, the 5% quantiles for each herd are depicted in Figure 3.3. There were a lot of differences between the curves in case of *C. bovis* while differences between the herd-specific curves of *S. aureus* and *S. uberis* are less pronounced. This suggests that there are substantial differences between herds for time to infection with *C. bovis* while there are less differences for time to infection with *S. aureus* or *S. uberis*.

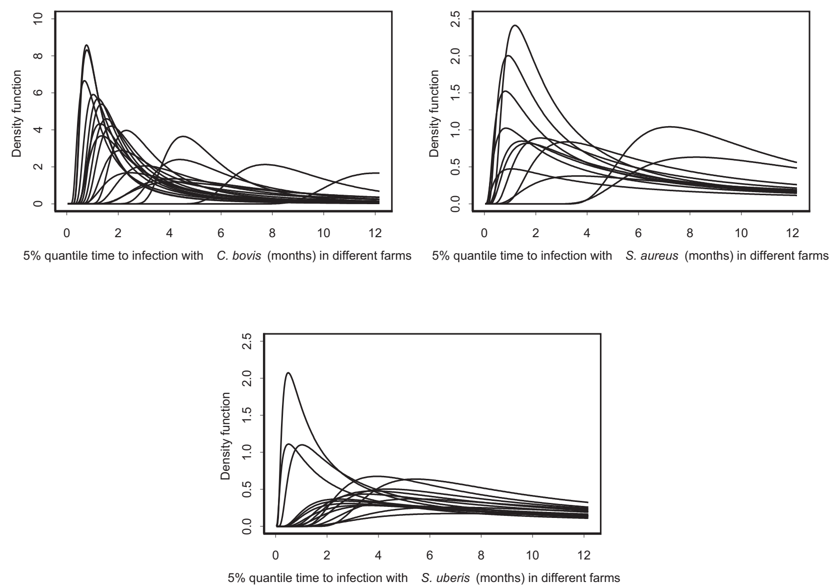

Figure 3.3: The 5% quantiles of the first infection (in months) for the different herds in the stratified analysis. Each line corresponds to a different farm

A likelihood ratio test was performed to test if the stratified model was better compared to the unstratified model, fitted on the same subset of herds. In the case of *C. bovis* ($P < 0.001$) and *S. uberis* ($P < 0.001$) there where significant differences between the baselines of the different herds. In case of *S. aureus* no significant differences could be observed ($P = 0.25$).

3.4 Discussion

In this paper, a method based on the gamma frailty model is proposed which handles clustering and interval censoring in an analytic way. Bellamy et al. (2004) also proposed a technique but rather assuming lognormally distributed frailties. In their proposal however, the random effect has to be integrated out numerically.

The substantial clustering of infection times within the cow can have different reasons. First of all there are a lot of cow-associated factors: the physiological status and the environment is shared among quarters. It is also possible that during the milking process, teats are infected by the milking machine from one quarter to another. The high clustering can also be the result of a confounder at cow level (e.g. breed and teat confirmation (de Haas et al., 2003; Neijenhuis et al., 2001)). Probably a part of the clustering can be explained by unknown factors at the herd level (hygiene level, outdoor pasture, teat dipping or disinfection (Lam et al., 1997), production level etc.) or in herd cow-to-cow spread during the milking process (Zadoks et al., 2003). Nash et al. (2003) also demonstrated a genetic influence on susceptibility to IMI. High intra-cow clustering is also found by Schukken et al. (1999) and Moret-Stalder et al. (2009). A hierarchical model that can estimate the variance of frailties on different levels could estimate the variance on herd level, but those models are not yet available for interval censored data.

The reduction of the variance of the frailties in the stratified model suggests that a considerable part of the variation between the cows in the unstratified model can be explained by the differences between herds. The differences in baseline hazards between herds were highest for *C. bovis*, smaller for *S. uberis* and not significant for *S. aureus*. Barkema et al. (1997) also found high clustering at herd level for *C. bovis* which is an indicator for differences between herds. The cow-to-cow transmission can be reduced by good hygiene and management practices and not all herds have the same hygiene level. This results in large differences between herds. *S. uberis*, on the other hand, is in most cases known as an environmental pathogen. How-

ever, in some herds there are also infectious strains (Zadoks et al., 2001a; McDougall et al., 2004). The variance of the frailties on cow level within the same herd remains high for all three bacteria, especially for *S. uberis*, probably due to differences in susceptibility of individual cows.

Ignoring the interval censoring results in a less accurate estimation of the baseline hazard which is important when one wants to compare the hazard functions. The problem of interval censoring is sometimes solved using the midpoint between the last negative and first positive sample (Zadoks et al., 2003; McDougall et al., 2004). Using the midpoint leads in our dataset to a slight underestimation of the hazard in the first period. A possible explanation for the underestimation of the scale and overestimation of the shape-parameter, is the lack of information at the beginning of the study (Table 3.1). If only the midpoint is taken into account, no events occur in the first weeks, which results in a lower event rate in the first weeks and a steper increase of the hazard after this initial period. This effect is even more pronounced if the upper bound is used in stead of the midpoint (results not shown). This results in a lower baseline hazard and, as a result of this, in our analysis, the naive model slighty overestimated the fixed effect. Radke (2003) demonstrated that ignoring asynchronous interval censoring results in most cases in overestimation of the risk factor regression coefficients in their absolute size to even more than 50%. The results of the proposed model still indicates an increasing hazard for intramammary infection during lactation, although not significant anymore for *S. uberis* and *S. aureus*. Gasqui et al. (2000) also found an increasing hazard for mastitis during one lactation.

The time to IMI for *C. bovis* and *S. uberis* is affected by parity. Multiparous cows are more likely to get infected earlier in lactation than primiparous cows. This effect is also mentioned by other authors (Zadoks et al., 2001b). The time to IMI with *S. aureus* is not affected by parity (p=0.20). Some authors have the same conclusion (Pitkälä, 2004; Schukken et al., 1999) while others claim the opposite (Zadoks et al., 2001b; Moret-Stalder et al., 2009).

3.5 Conclusion

The use of a model that takes into account interval censoring and clustering gives more accurate estimates and additionally a measure for clustering. The proposed model gives, especially for the estimation of the baseline hazard,

better results than a model that ignores interval censoring. The estimation of the frailty variance and the fixed effect seems to be less affected by interval censoring. When the purpose of the analysis is to estimate the hazard functions then a model that takes into account interval censoring is definitely preferred.

Multiparous cows were more likely to get infected earlier in lactation with *C. bovis* or *S. uberis* than primiparous cows while the time to intramammary infection with *Staphylococcus aureus* is not significantly influenced by parity.

All three bacteria are highly clustered on cow level within the population and a stratified analysis, on a subset of herds, also suggests a substantial clustering on herd level for *C. bovis* and *S. uberis*. The clustering on cow level within the herd seems te be highest in case of *S. uberis*.

Bibliography

Allore H.G. and Erb H.N., 1999. Approaches to modeling intramammary infections in dairy cattle. Prev. Vet. Med. 39, 279-293.

Barkema H.W., Schukken Y.H., Lam T.J.G.M., Galligan D.T., Beiboer M.L. and Brand A., 1997. Estimation of interdependence among quarters of the bovine udder with subclinical mastitis and implications for analysis. J. Dairy Sci. 80, 1592-1599.

Bellamy S.L., Li Y., Ryan L.M., Lipsitz S., Canner, M.J. and Wright R., 2004. Analysis of clustered and interval censored data from a community-based study in asthma. Stat. Med. 23, 3607-3621.

Bradley A.J., 2002. Bovine mastitis: an evolving disease. Vet. J. 164, 116-128.

Collet D., 2003. Modelling survival data in medical research 2nd ed. Chapman & Hall, Boca Raton.

Compton C.W.R., Heuer C., Parker K. and McDougall S., 2007. Risk factors for peripartum mastitis in pasture-grazed dairy heifers. J. Dairy Sci. 90, 4171-4180.

de Haas Y., Barkema H.W., Schukken Y.H. and Veerkamp R.F., 2003. Genetic associations for pathogen-specific clinical mastitis and patterns of peaks in somatic cell count. Anim. Sci. 77, 187-195.

Duchateau L. and Janssen P., 2005. Understanding heterogeneity in generalized mixed and frailty models. Am. Stat. 59, 143-146.

Duchateau L. and Janssen P., 2008. The Frailty Model. Springer, New York.

Duchateau L., Janssen P., Lindsey P., Legrand C., Nguti R. and Sylvester R., 2002. The shared frailty model and the power for heterogeneity tests in multicenter trials. Comput. Stat. Data An. 40, 603-620.

Gasqui P., Pons O. and Coulon J-B., 2000. An individual modelling tool for consecutive clinical mastitis during the same lactation in dairy based on a survival model. Vet. Res. 31, 583-602.

Glidden D.V. and Vittinghoff E., 2004. Modelling clustered survival data from multicentre clinical trials. Statist. Med. 23, 369-388.

Goethals K., Ampe B., Berkvens D., Laevens H., Janssen P. and Duchateau L., 2009. Modeling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. J. Agric. Biol. Environ. Stat. 14, 1-14.

Klein J.P. and Moeschberger M.L., 1997. Survival analysis. Techniques for censored and truncated data. Springer Verlag, New York.

Laevens H., Deluyker H., Schukken Y. H., De Meulemeester L., Vandermeersch R., De Muelenaere E. and De Kruif A., 1997. Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. J. Dairy Sci. 80, 3219-3226.

Lam T.J.G.M., van Vliet J.H., Schukken Y.H., Grommers F.J., van Velden-Russcher A., Barkema H.W. and Brand A., 1997. The effect of discontinuation of postmilking teat disinfection in low somatic cell count herds. II Dynamics of intramammary infections. Vet. Q. 19, 47-53.

McDougall S., Parkinson T.J., Leyland M. Anniss M.F. and Fenwick S.G., 2004. Duration of infection and strain variation in *Streptococcus uberis* isolated from cows' milk. J. Dairy Sci. 87, 2062-2072.

Moret-Stalder S., Fournier C., Miserez R, Albini S., Doherr M.G., Reist M., Schaeren W., Kirchhofer M., Graber H.U., Steiner A. and Kaufmann T., 2009. Prevalence study of *Staphylococcus aureus* in quarter milk samples of dairy cows in the Canton of Bern, Switzerland. Prev. Vet. Med. 88, 72-76.

Nash D.L., Rogers G.W., Cooper J.B., Hargrove G.L. and Keown J.F., 2003. Heritability of intramammary infections at first parturition and relationships with sire transmitting abilities for somatic cell score, udder type traits, productive life, and protein yield. J. Dairy Sci. 86, 2684-2695.

Neijenhuis F., Barkema H.W., Hogeveen H. and Noordhuizen J.P.T.M., 2001. Relationship between teat-end callosity and occurrence of clinical mastitis. J. Dairy Sci. 84, 2664-2672.

Oakes D., 1989. Bivariate survival models induced by frailties. J. Am. Stat. Assoc. 84, 487-493.

Pitkälä A., Haveri M., Pyörälä S., Myllys V. and Honkanen-Buzalski T., 2004. Bovine mastitis in finland 2001 - prevalence, distribution of bacteria, and antimicrobial resistance. J. Dairy Sci. 87, 2433-2441.

Radke B.R., 2003. A demonstration of interval-censored survival analysis. Prev. Vet. Med. 59, 241-256.

Seegers H., Fourichon C and Beaudeau F., 2003. Production effects related to mastitis and mastitis economics in diary cattle herds. Vet. Res. 34, 475-491.

Schukken Y.H., Leslie K.E., Barnum D.A., Mallard, B.A., Lumsden J.H., Dick P.C., Vessie G.H. and Kehrli M.E., 1999. Experimental *Staphylococcus aureus* intramammary challenge in late lactation dairy cows: quarter and cow effects determining the probability of infection. J. Dairy Sci. 82, 2393-2401.

Zadoks R.N., Allore H.G., Barkema H.W., Sampimon O.C., Gröhn Y.T. and Schukken Y.H., 2001a. Analysis of an outbreak of *Streptococcus uberis* mastitis. J. Dairy Sci. 84, 590-599.

Zadoks R.N., Allore H.G., Barkema H.W., Sampimon O.C., Wellenberg G.J., Gröhn Y.T. and Schukken Y.H., 2001b. Cow- and quarter-level risk factors for *Streptococcus uberis* and *Staphylococcus aureus* mastitis. J. Dairy Sci. 84, 2649-2663.

Zadoks R.N., Allore H.G., Hagenaars T.J., Barkema H.W. and Schukken Y.H., 2002. A mathematical model of *Staphylococcus aureus* control in dairy herds. Epidemiol. Infect. 129, 397-416.

Zadoks R.N., Gillespie B.E., Barkema H.W., Sampimon O.C., Oliver S.P. and Schukken Y.H., 2003. Clinical, epidemiological and molecular characteristics of *Streptococcus uberis* infections in dairy herds. Epidemiol. Infect. 130, 335-349.

Chapter 4

Investigating the protective effect of coagulase-negative staphylococci or *Corynebacterium bovis* infection on intramammary pathogenic infections using a frailty model based approach.

Based on:
Ampe B., Goethals K., Laevens H., Burvenich C. and Duchateau L., "Investigating the protective effect of coagulase-negative staphylococci (CNS) or *Corynebacterium bovis* infection on intramammary pathogenic infections using a frailty model based approach." *in preparation.*

Abstract

This study examines the influence of intramammary infections (IMI) with coagulase-negative staphylococci (CNS) or *Corynebacterium bovis* on the hazard of infection with other mastitis pathogens.

In a longitudinal observational study with in total 1132 cows on 25 dairy herds, the time to infection with other mastitis pathogens was measured based on monthly milk samples at quarter level. This results in interval-censored time to event data. The infection status with CNS and *C. bovis*, a time varying covariate, was also registered. Frailty model methodology was used to analyze the time to event data taking into account clustering (quarters in udders), interval censoring and the time varying CNS or *C. bovis* infection status. The presented model was validated in a separate simulation study based on 500 datasets and suggested that the proposed technique was appropriate to analyze the data.

Quarters infected with CNS were more vulnerable to an IMI with *S. uberis* (HR: 2.02, P = 0.005) and *S. dysgalactiae* (HR: 3.29, P < 0.001) than uninfected quarters. On the other hand, an IMI with *C. bovis* did have a significant protective effect against an IMI with *S. aureus* (HR: 0.54, P = 0.002), *S. uberis* (HR: 0.22, P < 0.001) or *S. dysgalactiae* (HR: 0.32, P = 0.004). No significant effect of an IMI with *C. bovis* on the occurrence of *E. coli* and other coliform bacteria was found (P > 0.37).

4.1 Introduction

Coagulase-negative Staphylococci (CNS) are the most prevalent bacteria in bovine milk samples in many countries (Pyörälä and Taponen, 2009). Piessens et al. (2012) investigated the most common species of the coagulase-negative Stapylococci in six Flemish dairy farms. The most common species were *S. haemolyticus*, *S. simulans*, *S. chromogenes* and *S. epidermidis*, each with a number of different genotypes.

It is generally accepted that IMI caused by CNS do not induce clinical mastitis. Some species induce a weak inflammatory response, with a mild increase of the SCC. Most CNS species cause a persistent IMI (Supré et al., 2011). Piepers et al. (2010) indicate that an IMI with CNS in early lactation results in a higher daily milk yield compared to noninfected heifers. Heifers infected with CNS have less often clinical mastitis during the following lactation. Schukken et al. (2009) also found a slightly higher milk production in CNS infected cows.

Matthews et al. (1991) suggest that quarters harboring CNS suppress colonization of the mammary gland by major mastitis pathogens (*S. aureus*, *S. dysgalactiae*, coliforms, *S. uberis* and other streptococci). The study compares the infection rates of new IMI with major pathogens in quarters with or without pre-existent CNS infections. A chi-square test of independence was used to test the difference, without correction for clustering of quarters within cows. Michel et al. (2011) reports a lower prevalence of CNS in herds with high prevalence of *S. aureus*.

Also IMI caused by *C. bovis* are generally accepted not to induce clinical mastitis. Some species induce a weak inflammatory response, with only a mild increase of the SCC. Most *C. bovis* strains colonize the teat duct and cause a persistent IMI (Brooks and Barnum, 1984a). Infections with *C. bovis* receive little attention in Mastitis research.

Pankey et al. (1985) and Hogan et al. (1988) found a protective effect of *C. bovis* infection on the susceptibility to some major mastitis pathogens but also the opposite for some others. Brooks et al. (1983) did not found a protective effect in an observational study but did found a protective effect against *S. aureus* in an experimental study where control quarters and quarters colonized with C.bovis where challenged by inoculation with *S. aureus* (Brooks and Barnum, 1984b).

In previous studies the analysis was mostly based on the analysis of binary data, infected versus uninfected as cows were only evaluated once. Further, the analysis is sometimes based on quarter level data (Schukken et al., 1999) with proper adjustment for clustering while others do not correct for clustering (Matthews et al., 1991; Davidson et al., 1992). Barkema et al. (1997) discussed the possible implications of ignoring clustering, which can result in the underestimation of the variance of the estimates and the associated increase of the Type 1 error.

The aim of this study was to investigate the effect of IMI with CNS or *C. bovis* on the hazard of infection with other mastitis-related pathogens under field conditions, using a novel frailty model methodology to take into account all data aspects. In previous mentioned studies, quarters were considered as infected or not infected without an evolution over time and only evaluated after a predefined time. A model with time varying covariates makes it possible to follow a quarter over time. In most cases a quarter starts uninfected followed by some infection cycles with (persistent) CNS. The hazard of infection for each quarter changes over time according to the infection status of that quarter. If an IMI with CNS or *C. bovis* has an effect on the hazard of infection with another mastitis pathogen, the estimated effect will be significantly negative (hazard ratio smaller than 1), which results

in a lower hazard to infection.

We focus on, *S. aureus*, *S. uberis*, *S. dysgalactiae*, *E. coli* and other coliform bacteria. The frailty model used in Ampe et al. (2012) takes clustering and interval censoring into account. In this study, the frailty model methodology is extended to deal with time varying covariates which make it possible to include the infection status with CNS and *C. bovis* at any time in lactation.

4.2 Materials and methods

4.2.1 Intramammary Infections Dataset

In total, 1132 cows on 25 dairy herds, located in the provinces East and West Flanders, Belgium, were followed during a 20-months period (February 1993 to September 1994). Only lactations with parturition after February 1993 were considered for analysis.

At monthly intervals (11 times a year, not during either July or August) quarter foremilk samples were taken to detect intramammary infections (IMI).The teats were cleaned with dry udder cloths. Dirty teats were washed and dried. Before milk samples were taken, all teats were disinfected with cotton moistened with a solution of ethyl alcohol (70%) and chlorhexidine (200 mg/100 ml). Immediately after collection, the milk samples were transported to a laboratory and 0.01 ml aliquots were streaked for initial isolation within 2 to 3 h after collection onto a 90-mm Petri dish with a blood agar base (Oxoid, Basingstoke, England) supplemented with 5% bovine blood. Samples were also streaked onto an Edwards medium (Oxoid) supplemented with 5% bovine blood. Agar plates were incubated at 37°C and read after 24 and 48 h (Laevens et al., 1997). A quarter was considered to be positive when more than 1 cfu of the considered species was found in a 0.01 mL aliquot. Dohoo et al. (2011) suggested this definition as an acceptable definition for most species.

Different bacteria were isolated and identified as described by the National Mastitis Council. In this paper we focus on *S. aureus*, *S. uberis*, *S. dysgalactiae*, *E. coli*, other coliform bacteria, *C. bovis* and coagulase-negative staphylococci (CNS). The other isolated bacteria were too uncommon for further analysis. More details about the isolation and identification procedure are given by (Laevens et al., 1997). A quarter was assumed to be infected with CNS starting from the month before the observed isolation of CNS or *C. bovis*. As long as the germ was isolated during the following intervals, the quarter was assumed to be infected during the preceding

interval. Dohoo et al. (2011) showed that all culture procedures based on one milk sample, have a limited sensitivity, particularly in case of CNS and Streptococcus spp. Therefore, a negative sample preceded and followed by a positive sample was assumed to be a false negative value and converted to a positive value.

The time to first IMI with a major pathogen was determined for each udder quarter and constitutes the response variable. Because previous research (Ampe et al., 2012) showed a significant effect of parity (primiparous versus multiparous) for some bacteria, parity was also included as a fixed effect.

4.2.2 Data Analysis

Individual udder quarter infection times can be modeled most efficiently by survival analysis techniques, because not in all quarters an infection occurred before the end of the study which results in censoring. For quarter j ($j = 1, 2, 3, 4$) from cow i ($i = 1, \ldots, n$) the true infection time is only known to be between the last observed time without infection (lower bound, L_{ij}) and the first time with infection (upper bound, U_{ij}) for interval censored quarters and after time R_{ij} for right censored quarters, i.e. no infection occurred during the follow up period of the quarter. The dataset considered in this paper has specific characteristics, a time varying covariate, requiring an extension of the shared frailty model for interval censored data proposed by Ampe et al. (2012).

The complete risk time is split up in periods delineated by the sampling times l_{ijk} ($k = 1, , n_{ij}$ with $l_{ij0} = 0$), as the CNS infection status might change at such times. For period k, delineated by sampling times l_{ijk-1} and l_{ijk}, the covariate corresponds to x_{ijk}, taking value zero or one (not infected or infected).

The modeling is based on a parametric proportional hazards shared frailty model with Weibull baseline hazard with scale parameter λ and shape parameter γ. In this paper the parameters of interest were the fixed effects parity, expressed by parameter β_p, and the effect of CNS or *C. bovis* infection, expressed by parameter β_{CNS} or β_{CBO}. Parity (primiparous or multiparous) is a cow variable, and in the time scope of this study constant over time and denoted by x_i for cow i.

Finally, we introduce a frailty for cow, u_i , assumed to come from a one-parameter gamma density with mean one and variance θ, to adjust for clustering of the quarters in the cow. This leads to the following conditional

hazard model expression:

$$h_{ij}(t) = \begin{cases} \lambda\gamma t^{\gamma-1} u_i \exp(\beta_{cns} x_{ij1} + \beta_p x_i) & \text{for} \quad l_{ij0} = 0 < t < l_{ij1} \\ \vdots \\ \lambda\gamma t^{\gamma-1} u_i \exp(\beta_{cns} x_{ijk} + \beta_p x_i) & \text{for} \quad l_{ijk-1} < t < l_{ijk} \\ \vdots \\ \lambda\gamma t^{\gamma-1} u_i \exp(\beta_{cns} x_{ijn_{ij}} + \beta_p x_i) & \text{for} \quad l_{ijn_{ij}-1} < t < l_{ijn_{ij}} \end{cases}$$

which results in the conditional cumulative hazard expression:

$$H_{ij}(t) = \sum_{k=1}^{l_{ijk} \leqslant t} u_i \lambda (l_{ijk}^{\gamma} - l_{ijk-1}^{\gamma}) \exp(\beta_{cns} x_{ijk} + \beta_p x_i)$$

from which the conditional survival function can be obtained as:

$$S_{ij}(t) = \exp(-H_{ij}(t))$$

The conditional likelihood for data with both right and interval censoring is given by (Collet, 2003; Duchateau and Janssen, 2008):

$$L = \prod_{i=1}^{n} \prod_{j=1}^{4} (S_{ij}(L_{ij}) - S_{ij}(U_{ij}))^{\delta_{ij}} (S_{ij}(R_{ij}))^{(1-\delta_{ij})}$$

The censor indicater (δ_{ij}) takes value 0 in case of right censoring and 1 in case of interval censoring. Inserting the relevant expression for the survival function leads to the conditional likelihood for cow i:

$$\begin{aligned} L(\lambda, \gamma, \boldsymbol{\beta} | \boldsymbol{u}) &= \prod_{i=1}^{n} \prod_{j=1}^{4} \Big((\exp(-H_{ij}(L_{ij})) - \exp(-H_{ij}(U_{ij})))^{\delta_{ij}} \\ &\quad \times (\exp(-H_{ij}(R_{ij})))^{(1-\delta_{ij})} \Big) \end{aligned}$$

This conditional likelihood contains the unobserved frailty term u_i. We can integrate out the frailty using its distributional assumption:

$$\begin{aligned} L(\theta, \lambda, \gamma, \boldsymbol{\beta}) &= \prod_{i=1}^{n} \int_0^{\infty} \prod_{j=1}^{4} \Big((\exp(-H_{ij}(L_{ij})) - \exp(-H_{ij}(U_{ij})))^{\delta_{ij}} \\ &\quad \times (\exp(-H_{ij}(R_{ij})))^{(1-\delta_{ij})} \Big) \frac{u_i^{1/\theta-1} \exp(-u_i/\theta)}{\theta^{1/\theta} \Gamma(1/\theta)} du \end{aligned}$$

Which can be done analytically. The obtained marginal likelihood does no longer contain the unobserved frailties, but only the variance (θ) of the frailties, and can be maximized by general purpose maximizers, such as the nlm function in R. The nlm function also estimates the hessian matrix at the maximum. The inverse of the hessian matrix is an estimate of the variance covariance matrix of the estimated parameters ($\theta, \lambda, \gamma, \boldsymbol{\beta}$).

The estimates for the fixed effects are commonly transformed to conditional hazard ratios ($HR = \exp(\beta)$), conditional on the same frailty.

The variance of the frailties was transformed to the non-parametric intraclass correlation coefficient, Kendall's τ, given by $\theta/(2 + \theta)$ (Glidden and Vittinghoff, 2004). Kendall's τ is a global measure of dependence that gives an idea about the correlation between the four quarters within the same cow. The interpretation of the variance of the frailties is discussed more in detail in Duchateau and Janssen (2005).

4.2.3 Simulation Study

To validate the proposed model, a small simulation study was performed. 500 data sets for each setting of parameters were generated. Each data set contained 1000 cows housed in 20 farms. A parity (primiparous or multiparous, x_i) was randomly assigned at cow level. A random frailty for each cow was generated from the one-parameter gamma distribution with variance θ. The value of a time varying covariate (x_{ijk}) changed over time at randomly generated time points, dividing the lactation into different periods.

In the next step a survival function was created for each quarter according to the scale and shape parameter (λ and γ) of the Weibull distribution, the value of the frailty and the covariates at each time point. A random value of the survival function (between 0 and 1) was created for each quarter with the runif() function in R. Using the inverse of the survival function, the corresponding event time was calculated. If the event time was lower than the end of the lactation, a random interval was created around the event time, in the other case the quarter was right censored. The choice of the values of the parameters for simulation was based on the values obtained in the analysis of the dataset in order to create datasets with a comparable number of events.

4.3 Results

4.3.1 Intramammary Infections Dataset

Quarter infection data were obtained from 4527 quarters of 1132 cows on 25 farms. An IMI with CNS occurred in 792 (17.5%) quarters, an IMI with *C. bovis* occurred in 1932 (42.7%) quarters. The number of infected quarters with the different bacteria was: *S. aureus*: 228 (5.0%), *S. uberis*: 175 (3.9%), *S. dysgalactiae*: 86 (1.9%), *E. coli*: 31 (0.7%) and the other coliform bacteria: 50 (1.1%). Sixty-six percent of the cows were multiparous.

The results of the analysis for time to IMI with the 5 different bacteria and the effect of lactation and an existing IMI with CNS or *C. bovis* are summarized in Table 4.1 and Table 4.2.

The variance of the frailties, θ, is due to the differences between the cows but also due to the differences between different farms. The variance of the frailties for the analysis with CNS as time varying covariate was high for *S. aureus* (5.76) and *S. uberis* (6.18) which results in a high Kendall's τ (0.74 and 076). The variance of the frailties of *S. dysgalactiae* (3.05 - Kendall's τ 0.60) and the variance of the frailties of *E. coli* (1.32 - Kendall's τ 0.39) and other coliform bacteria (3.34 - Kendall's τ 0.62) were lower and not significantly different from 0. The variances of the frailties for the analysis with time varying variable *C. bovis* were similar.

Parity had a significant effect on the hazard to have an IMI with *S. uberis* and *E. coli*, for both the estimate is 1.07 which corresponds with the respective conditional hazard ratios of 2.90 (95% CI: [1.80 ; 4.69]) and 2.93 (95% CI: [1.11 ; 7.72]) where multiparous cows had a higher hazard to have an IMI than primiparous cows. No other significant effect of parity was found for the other bacteria. The estimates based on the model with *C. bovis* as time varying covariate were similar.

Table 4.1: Estimates and standard error of the parameters θ (variance of the frailties), λ and γ (scale and shape parameters of the Weibull distribution) and the fixed effect for the variable parity (multiparous versus primiparous) and time varying variable infection with CNS (β_{CNS}, infected versus uninfected). The estimates are different from 0 (or 1 in case of γ) with $^{\dagger}P < 0.10$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

	θ	λ	γ	β_{parity}	β_{CNS}
S. aureus	5.71***± 0.95	0.08***± 0.01	1.12 ± 0.07	0.29 ± 0.21	-0.25 ± 0.30
S. uberis	5.43***± 1.09	0.03**± 0.01	1.11 ± 0.08	1.07***± 0.24	0.70**± 0.24
S. dysgalactiae	2.82*±* 1.34	0.02**± 0.01	1.00 ± 0.11	0.14 ± 0.26	1.19***± 0.31
Escherichia coli	1.06 ± 2.07	0.01 ± 0.00	1.47*± 0.26	1.07*± 0.49	0.08 ± 0.74
other coliform	3.03 ± 2.11	0.01**± 0.00	1.17 ± 0.16	0.56 ± 0.35	0.63 ± 0.48

Table 4.2: Estimates and s.e. of the parameters θ, λ and γ and the fixed effect for the variable parity (multiparous versus primiparous) and time varying variable infection with *C. bovis* (β_{CBO}, infected versus uninfected). The estimates are different from 0 (or 1 in case of γ) with $^{\dagger}P < 0.10$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

	θ	λ	γ	β_{parity}	β_{CBO}
S. aureus	5.97***± 0.98	0.09***± 0.02	1.17*± 0.07	0.35 ± 0.22	-0.62**± 0.20
S. uberis	5.45***± 1.07	0.04**± 0.01	1.21*± 0.09	1.14***± 0.25	-1.50***± 0.32
S. dysgalactiae	3.42*± 1.51	0.03**± 0.01	1.10 ± 0.11	0.19 ± 0.26	-1.13**± 0.38
Escherichia coli	0.92 ± 1.91	0.01 ± 0.01	1.53*± 0.26	1.10*± 0.55	-0.42 ± 0.47
other coliform	2.88 ± 2.05	0.01 ± 0.01	1.21 ± 0.17	0.57 ± 0.36	-0.34 ± 0.38

The effect of an IMI with CNS on the hazard to have an IMI with other bacteria was only significant in case of *S. uberis* (0.70) and *S. dysgalactiae* (1.19). This corresponded to a HR of 2.02 (95% CI: [1.25 ; 3.27]) in case of *S. uberis* and a HR of 3.29 (95% CI: [1.80 ; 6.01]) in case of *S. dysgalactiae*. This means that an IMI with CNS increased the hazard to have a new IMI with *S. uberis* and *S. dysgalactiae*. For *E. coli* (0.08, P=0.92) no effect was found and the other coliform bacteria (0.63, P=0.19) a not significant positive effect was found while for *S. aureus* a not significant negative effect was found (-0.25, P=0.39).

The effect of an IMI with *C. bovis* on the hazard to have an IMI with other bacteria was significantly negative in case of *S. aureus* with a corresponding HR of 0.54 (95% CI: [0.36 ; 0.80]), *S. uberis* with a corresponding HR of 0.22 (95% CI: [0.12 ; 0.42]) and *S. dysgalactiae* with a corresponding HR of 0.32 (95% CI: [0.15 ; 0.68]). The effect of an IMI with *C. bovis* on the hazard to have an IMI with *E. coli* or other coliform bacteria is also negative but the effect was not significant (P=0.37).

The interaction between infection with CNS and lactation and infection with *C. bovis* and lactation was also tested in both models (results not shown). In case of infection with CNS the estimated effect for the interaction was slightly negative for all bacteria, but none of them was significantly different from 0 (all P-values > 0.2). In case of infection with *C. bovis* the estimated effect was slightly positive for all bacteria but also not significant (all P-values > 0.69). The interaction term was deleted from the final models.

Due to convergence problems, the timescale was rescaled to years instead of days, so the scale parameter l has to be interpreted on a year scale.

4.3.2 Simulation Study

The results of the simulation study are summarized in Table 4.3. The proposed technique was able to estimate the parameters of interest with a coverage around 95% for all estimated parameters. Also the variance of the frailties and the estimates of the Weibull parameters were estimated very well. The power to detect an existing difference was high for high differences but decreased when the effect size became smaller. The power to detect a time varying effect of 0.7 (as in simulation 1) was almost 100% while the power to detect a real effect of 0.3 in a similar setting dropped to 34% (data not shown). Also a reduction of the number of events, as in simulation 2 with a lower lambda, results in a lower power to detect differences. The results of the simulation study suggest that the proposed technique analyses

Table 4.3: Results of the analysis of simulated data sets using the proposed model.

Parameter	True value	Mean Estimate	Coverage (%)	Power (%)
Simulation 1:				
θ	5	5.054	95.4	
λ	0.08	0.081	94.6	
γ	1.1	1.126	93.2	
β_{parity}	0.7	0.721	93.6	95.6
β_{CNS}	0.7	0.680	95.6	100
Simulation 2:				
θ	3	2.945	92.6	
λ	0.01	0.010	94.4	
γ	1.1	1.142	96.8	
β_{parity}	0.56	0.609	96.6	34.2
β_{CNS}	0.63	0. 592	95.4	48.6

the data set in a proper way. The coverage of the parameters of interest was close to 95% and the average of the estimates was close to the real values.

4.4 Discussion

The hazard for a new IMI with *S. uberis* or *E. coli* was higher for multiparous cows compared to primiparous cows. The hazard for a new IMI with *S. aureus*, *S. dysgalactiae* and other coliform bacteria was not significantly influenced by parity.

For *S. aureus*, no significant effect of an IMI with CNS was observed. The effect of CNS was slightly negative (-0.26) but not significantly different from zero (P=0.40). It has been suggested that CNS have a protective effect (Pankey et al., 1985; Davidson et al., 1992; Michel et al., 2011) but a significant causal effect was not proved. De Vliegher et al. (2004) showed an inhibitory activity of some CNS strains, isolated from teat apices of heifers, against *S. aureus*, but this was based on in vitro experiments.

The hazard for a new IMI with *S. uberis* and *S. dysgalactiae* is increased when the quarter is infected with CNS. A possible explanation is that a cow who is sensitive to CNS IMI is also more sensitive to other environmental

bacteria. This is in contrast with other research (De Vliegher et al., 2004; Nascimento et al., 2005) that demonstrated the production of bacteriocins by some CNS strains against streptococcal bacteria. But not all CNS strains have the capacity to produce these bacteriocins. The effect of the bacteriocins was tested in vitro, while our data are observational field data without identification of species, no information on the prevalence and distribution of the different species was obtained.

On the other hand, a quarter that is more resistant against CNS is probably also more resistant against other pathogens. The differences in susceptibility are due to physiologic (parity, lactation stage,...) genetic (teat conformation) and environmental factors. An environment highly contaminated with one bacteria is probably also contaminated with other bacteria. Lam et al. (1997) also found a higher infection rate with *S. aureus* in quarters infected with coagulase-negative Micrococceae. In our opinion and based on the contradictory results in literature, the often claimed protective effect of CNS against major pathogens is less relevant in the mastitis control.

Both *E. coli* and the other coliform bacteria were not influenced by IMI with CNS. De Vliegher et al. (2004) also did not found an inhibitory effect of CNS (*S. chromogenes*) bacteriocins in an in vitro experiment. The inhibitory effect of bacteriocins is commonly more effective against related species and genera. The larger phylogenetical distant between the bacterial families was suggested as a possible explanation of the almost complete lack of inhibitory effect on the bacterial growth of the coliform bacteria. Also other authors did not found a protective effect of CNS on *E. coli* or other coliform bacteria (Hogan et al., 1988; Lam et al., 1997).

An infection with *C. bovis* did have a protective effect (HR: 0.54) against new IMI with *S. aureus*. Pankey et al. (1985) and Schukken et al. (1999) found similar results based on challenges. Hogan et al. (1987) showed that *S. aureus* and *C. bovis* are competitive for rate-limiting substrates in milk in vitro which results in a lower growth rate when *C. bovis* cultures were present in the milk.

The effect of an infection with *C. bovis* on the hazard for a new IMI with other bacteria was significantly negative for *S. aureus*, *S. uberis* and *S. dysgalactiae*. This suggests that an infection with *C. bovis* has a protective effect against those pathogens.

The protective effect of an IMI with *C. bovis* was not significant for *E. coli* and other coliform bacteria. Most authors did not find an effect of an IMI with minor pathogens on the hazard for a new IMI with *E. coli* or other coliform bacteria (Hogan et al., 1988; Lam et al., 1997).

4.5 Conclusions

The proposed frailty model is capable to analyze clustered interval censored survival data with time varying covariates. Multiparous cows are more vulnerable to an IMI with *S. uberis* or *E. coli* compared to primiparous cows.

Although a lot of authors claimed the protective effect of an IMI with CNS, our field study could not prove a protective effect against most mastitis pathogens, on the contrary, quarters infected with CNS had a higher hazard to have an IMI with *S. uberis* or *S. dysgalactiae*.

An IMI with *C. bovis* did result in a lower hazard to have an IMI with *S. aureus*, *S. uberis* or *S. dysgalactiae* but now effect was observed for *E. coli* or other coliform bacteria. These results suggest that an intra-mammary infection with *C. bovis* has a protective effect against some mastitis pathogens.

Appendix

The models in this chapter and in Chapter 3 can be fitted using the code in this section. The functions make use of a dataset with the structure as explained in Table 4.4.

Table 4.4: Structure of the IMI dataset.

Nr	Variable name	Explanation
1	cluster	cow number
2	id	quarter number
3	start	start of interval for timevarying covariate, this is always equal to 0 if no timevarying covariates are included.
4	left	lower border for an interval censored observation
5	time	right censoring time or upper border for an interval censored observation
6	fail	event indicator during that period: 0 for right censored, 1 for interval censored
7	failtot	overall event indicator: 0 for right censored, 1 if there is an event in that quarter over the full lactation = sum of 'fail' during all periods for that quarter
8 ...	variables ...	the variables in the model

The marginal loglikelihood has to be maximized. Because in R maximization is not implemented, we use a function to minimize the negative version of the loglikelihood. If code 1 is obtained, there was convergence. If after 20 attempts no convergence was observed the program stops and in case of almost convergence (code=3) those results are printed to the screen.

```
for(i in 1:20)
{
print(t <- nlm(CalcLogLik,t$estimate,print.level=1,
               hessian=TRUE,exponential=0))
if(t$code==1) {para <- t$estimate;break}
}
if(t$code==3) para <- t$estimate

solve(t$hessian)
covmatr<-solve(t$hessian)
sterr<-c(0,0)
para2<-para
para2[3]<-para[3]-1

# section for presentation of the results
for(i in 1:length(para))
{
if(i==1)tedrukken<-
    "\n\n      \t coef        \t se(coef)    \t z      \t p\n"
sterr[i]<-sqrt(covmatr[i,i])
tedrukken<-paste(tedrukken
       ,parname[i],"   \t",round(para[i],digits=6),"   \t"
 ,round(sterr[i],digits=6),"    \t"
 ,round(para[i]/sterr[i],digits=6),"    \t"
 ,round(2*(1-pnorm(abs(para2[i]/sterr[i]))),digits=6),"\n")
}
cat(tedrukken)
```

The function ”CalcLogLik” calculates the loglikelihood for the full dataset, which is a sum of the loglikelihoods of the different clusters, calculated by the function ”CalcLogLikClust”. In some cases, due to convergence problems, it is better to use the exponential of θ, λ and γ to be sure that they are positive. In latter case the optional parameter exponential has to be set to ”TRUE”. There is also a list ”signs” needed for the signs of the crossproducts.

```
CalcLogLik <- function(x,exponential=FALSE)
{ if (exponential){x[1:3]<-exp(x[1:3])}
  # return negative value of log likelihood
  -sum(by(datasetint,datasetint$cluster,CalcLogLikClust,p=x))
}
# p[1] = theta
# p[2] = lambda
# p[3] = gamma
# p[4,...] = beta's

CalcLogLikClust <-function(data,p)
{
log(1/(p[1]^(1/p[1]))*(t(1/(CalcHt(data[data$failtot==0,],p)
+1/p[1]+log(Crossproduct(data[data$failtot==1,],p=p)))
^(1/p[1]))%*%signs[[nevents+1]]))
}

signs<-list(1,c(1,-1))
for(i in 3:10) signs[[i]]<-kronecker(signs[[i-1]],c(1,-1))
```

For the calculation of the loglikelihood for each cluster, a function for the cumulative hazard "CalcHT" and for the crossproduct "Crossproduct" is needed:

```
CalcHt <- function(data,p,right=1)
{
if(length(data[,1])==0)
   { return(0)} # return 0 if no right censoring or no events
if(right==0) # if lower bound
   return(sum(p[2]*(data$time[data$fail==0]^p[3]
   -data$start[data$fail==0]^p[3])
   *exp(as.matrix(data[data$fail==0,8:(4+length(p))])
   %*%as.matrix(p[4:length(p)])))+p[2]
   *(data$left[data$fail==1]^p[3]-data$start[data$fail==1]^p[3])
   *exp(as.matrix(data[data$fail==1,8:(4+length(p))])
   %*%as.matrix(p[4:length(p)])))
#perform if right==1, no value returned yet
   return(sum(p[2]*(data$time^p[3]-data$start^p[3])
   *exp(as.matrix(data[,8:(4+length(p))])
   %*%as.matrix(p[4:length(p)]))))
}
```

```
Crossproduct <- function(data,p)
{
nevents <<- sum(data$fail)
if(nevents==0) {return (1)}
intRster <- unlist(by(data,data$id,CalcHt,p=p))
intLster <- unlist(by(data,data$id,CalcHt,p=p,right=0))
crossprod<-c(exp(intLster[1]),exp(intRster[1]))
if(nevents>1)
{
if(nevents>10)
   cat("Too many events: ", nevents, " !You're in trouble!")
        for(ik in 2:nevents)
{
crossprod <-
kronecker(crossprod,c(exp(intLster[ik]),exp(intRster[ik])));
}
}
return(crossprod)
}
```

Bibliography

Ampe B., Goethals K., Laevens H. and Duchateau L., 2012. Investigating clustering in interval-censored udder quarter infection times in dairy cows using a gamma frailty model. Prev. Vet. Med. In press: http://dx.doi.org/10.1016/j.prevetmed.2012.04.004.

Barkema H.W., Schukken Y.H., Lam T.J.G.M., Galligan D.T., Beiboer M.L. and Brand A., 1997. Estimation of interdependence among quarters of the bovine udder with subclinical mastitis and implications for analysis. J. Dairy Sci. 80, 1592-1599.

Brooks B.W. and Barnum D.A., 1984a. Experimental Colonization of the Bovine Teat Duct with *Corynebacterium bovis* and the Effect on Milk Somatic Cell Counts. Can. J. Comp. Med. 48, 141-145.

Brooks B.W. and Barnum D.A., 1984b. The Susceptibility of Bovine Udder Quarters Colonized with *Corynebacterium bovis* to Experimental Infection with *Staphylococcus aureus* or *Streptococcus agalactiae*. Can. J. Comp. Med. 48, 146-150.

Brooks B.W., Barnum D.A. and Meek A.H., 1983. An Observational Study of *Corynebacterium bovis* in selected Ontario Dairy Herds. Can. J. Comp. Med. 47, 73-78.

Collet D., 2003. Modelling survival data in medical research 2nd ed. Chapman & Hall, Boca Raton.

Davidson T.J., Dohoo I.R., Donald A.W., Hariharan H. and Collins K., 1992. A cohort study of coagulase negative staphylococcal mastitis in selected dairy herds in Prince-Edward-Island. Can. J. Vet. Res. 56, 275-280.

De Vliegher S., Opsomer G., Vanrolleghem A., Devriese L.A., Sampimon O.C., Sol J., Barkema H.W., Haesebrouck F. and de Kruif A., 2004. In vitro growth inhibition of major mastitis pathogens by *Staphylococcus chromogenes* originating from teat apices of dairy heifers. Vet. Microbiol. 101, 215-221.

Dohoo I., Andersen S., Dingwell R., Hand K., Kelton D., Leslie K., Schukken Y. and Goddon S., 2011. Diagnosing intramammary infections: Comparison of multiple versus single quarter milk samples for the identification of intramammary infections in lactating dairy cows. J. Dairy Sci. 94, 5515-5522.

Duchateau L. and Janssen P., 2005. Understanding heterogeneity in generalized mixed and frailty models. Am. Stat. 59, 143-146.

Duchateau L. and Janssen P., 2008. The Frailty Model. Springer, New York.

Glidden D.V. and Vittinghoff E., 2004. Modelling clustered survival data from multicentre clinical trials. Statist. Med. 23, 369-388.

Hogan J.S., Pankey J.W. and Duthie A.H., 1987. Growth Responses of *Staphylococcus aureus* and *Streptococcus agalactiae* to *Corynebacterium bovis* Metabolites. J. Dairy Sci. 70, 1294-1301.

Hogan J.S., Smith K.L., Todhunter D.A. and Schoenberger P.S., 1988. Rate of Environmental Mastitis in Quarters Infected with *Corynebacterium bovis* and Staphylococcus Species. J. Dairy Sci. 71, 2520-2525.

Laevens H., Deluyker H., Schukken Y. H., De Meulemeester L., Vandermeersch R., De Muelenaere E. and De Kruif A., 1997. Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. J. Dairy Sci. 80, 3219-3226.

Lam T.J.G.M., van Vliet J.H., Schukken Y.H., Grommers F.J., van Velden-Russcher A., Barkema H.W. and Brand A., 1997. The effect of discontinuation of postmilking teat disinfection in low somatic cell count herds. II Dynamics of intramammary infections. Vet. Q. 19, 47-53.

Matthews K.R., Harmon R.J. and Langlois B.E., 1991. Effect of naturally occuring coagulase-negative Staphylococci infections on new infections by mastitis pathogens in the bovine. J. Dairy Sci. 74, 1855-1859.

Michel A., Syring C., Steiner A. and Graber H.U., 2011. Intramammary infections with the contagious *Staphylococcus aureus* genotype B in Swiss dairy cows are associated with low prevalence of coagulase-negative staphylococci and *Streptococcus spp.* Vet. J. 188, 313-317.

Nascimento J.D.S., Fagundes P.C., Brito M.A.V.D., dos Santos K.R.N. and Bastos M.D.D., 2005. Production of bacteriocins by coagulase-negative staphylococci involved in bovine mastitis. Vet. Microbiol. 106, 61-71.

Pankey J.W., Nickerson S.C., Boddie R.L. and Hogan J.S., 1985. Effects of *Corynebacterium bovis* infection on susceptibility to major mastitis pathogens. J. Dairy Sci. 68, 2684-2693.

Piepers S., Opsomer G., Barkema H.W., de Kruif A. and De Vliegher S., 2010. Heifers infected with coagulase-negative staphylococci in early lactation have fewer cases of clinical mastitis and higher milk production in their first lactation than noninfected heifers. J. Dairy Sci. 93, 2014-2024.

Piessens V., De Vliegher S., Verbist B., Braem G., Van Nuffel A., De Vuyst L., Heyndrickx M. and Van Coillie E., 2012. Intra-species diversity and epidemiology varies among coagulase-negative Staphylococcus species causing bovine intramammary infections. Vet. Microbiol. 155, 62-71.

Pyörälä S. and Taponen S., 2009. Coagulase-negative staphylococci-Emerging mastitis pathogens. Vet. Microbiol. 134, 3-8.

Schukken Y.H., Gonzalez R.N., Tikofsky L.L., Schulte H.F., Santisteban C.G., Welcome F.L., Bennett G.J., Zurakowski M.J. and Zadoks R.N., 2009. CNS mastitis: Nothing to worry about? Vet. Microbiol. 134, 9-14.

Schukken Y.H., Leslie K.E., Barnum D.A., Mallard B.A., Lumsden J.H., Dick P.C., Vessie G.H. and Kehrli M.E., 1999. Experimental *Staphylococcus aureus* Intramammary challenge in late lactation dairy cows: quarter and cow effects determining the probability of infection. J. Dairy Sci. 82, 2393-2401.

Supré K., Haesebrouck F., Zadoks R.N., Vaneechoutte M., Piepers S. and De Vliegher S., 2011. Some coagulase-negative staphylococcus species affect udder health more than others. J. Dairy Sci. 94, 2329-2340.

Chapter 5

General discussion and future research perspectives

In this section a general conclusion and recommendation for further research are formulated. Because this thesis contains both clinical and statistical topics, this chapter is divided in a clinical part and a statistical part.

5.1 Clinical part

In this thesis frailty models are illustrated using two datasets. This resulted in some relevant clinical results, but also created some new ideas for possible further research.

5.1.1 IBR dataset

The hyperimmunization study suggested that a hyperimmunization protocol using booster vaccinations, repeatedly administered each 6 months allows for a better control than the old IBR vaccination schemes. Compared to the non-intervention group (NIG), the two hyperimmunization groups, with repeated administration of attenuated and inactivated-virus gE-deleted marker vaccine (HIG 1) as well as inactivated vaccines (HIG 2), had a decreasing hazard to seroconvert compared to the NIG.

Since 2007, as an implementation of a Royal Decree (KB 22/11/2006 published in the law gazette on 04/01/2007, and adapted by KB 16/02/2011 - BS 25/02/2011) a voluntary eradication program with hyperimmunization with marker vaccines was started based on the results of the presented field study. By the end of 2009, only 574 herds officially performed hyperimmunisation (Jaarverslag DGZ, 2009). The real number of herds using hyperimmunisation was probably higher as a result of the permissive reporting. Since 2012 hyperimmunization is obliged in all herds.

In November 2010, only 118 herds (0.7%) were 'officially' seronegative for BoHV-1. In January 2012, already 3766 (23.5%) herds were 'officially' seronegative (Jaarverslag DGZ, 2011).The 'official' seronegative herds are herds that test negative for all cows older than 1 year. A yearly follow up sample is required to maintain the seronegative label.

Not all farms however performed such a test yet. The real number of seronegative herds is unfortunately unknown, because not all herds are tested and no data on the number of tested herds are available . If the incidence will further decrease, as suggested by the study, the seroprevalence at herd level will also decrease (due to natural turnover and removal of (older) infected cows).This will lead to an increase of the number of negative herds in the near future.

It would be worthwhile to investigate the effect of the obligation of hyperimmunization on the evolution of the incidence and seroprevalence. The reduced infection pressure on national level, due to the vaccine coverage of all herds, will have an impact on the incidence of the disease. Also the purchase of only IBR-seronegative cattle reduces the risk of new infections in the farm.

Another interesting topic for further research is to identify possible risk-factors associated with the failure to become or remain an official seronegative herd. Some of the data needed for such a study are probably already available in the animal healthcare association in Flanders ("Dierengezondheidszorg Vlaanderen").

The results of the analysis did not show a significant improvement due to the hyperimmunization protocols for the bulls in the mixed dairy-beef herds. On the other hand, a much higher clustering was observed for the bulls compared to the cows of the mixed dairy-beef herds. This indicates large differences between the herds and is probably a indication of an unobserved risk factor on farm level.

5.1.2 Intramammary infections dataset

The analysis of the IMI dataset suggested a high clustering of IMI within cow for *S. aureus*, *S. uberis*, *S. dysgalaciae* and *C. bovis*. Also the effect of parity and an IMI with *C. bovis* or CNS was investigated. Parity had a significant effect on the hazard of IMI with *S. uberis*, *C. bovis* and *E. coli*. Multiparous cows have a higher hazard to have an IMI than primiparous cows.

A significant protective effect of an IMI with *C. bovis* was found against *S. aureus*, *S. uberis* and *S. dysgalactiae*. No significant protective effect was found against *E. coli* or other coliform bacteria. These findings were in accordance with findings of other authors.

The effect of an infection with CNS is a different story. A quarter infected with CNS had a significantly increased hazard for a new IMI with *S. uberis* or *S. dysgalactiae*, while no other effects were found for the other mastitis pathogens. This is in contrast with a lot of other research.

The main problem in this discussion is the definition of CNS. Coagulase-negative staphylococci are a very heterogeneous group of bacteria. Cows are most likely infected with CNS from environmental sources but they can originate from different sources, including the milker's hands. Most IMI-causing CNS species are *S. haemolyticus*, *S. simulans*, *S. chromogenes* and *S. epidermidis*. Some of them cause persistent infections while others only

cause transient infections. The genetic diversity of some species is also very high (e.g. *S. haemolyticus*, *S. simulans*). Some of them are considered as commensal bacteria while others cause IMI (Piessens et al., 2012).

It is therefore difficult to draw a univocal conclusion. Everything depends on the composition of the CNS group in the IMI dataset. Some isolates were further typed to get the species. Unfortunately not all isolates were typed which make an analysis on species level difficult.

Further research should focus on the effect of the different CNS species separately. As shown by Piessens et al. (2012), the CNS group is very diverse and it is highly questionable if the use of the entity CNS in mastitis research is still an appropriate choice. It is probably better to divide the CNS group to smaller entities (or even to species level) according to their virulence factors. The protective effects in the published papers were mainly due to only one species or even one strain of a species. In an observational study, like ours, the majority without a protective effect can mask the protective effect of some other species.

Reyher et al. (2012) recently performed a meta-analysis of published studies on the effect of IMI with minor mastitis pathogens on the acquisition of new IMI with major mastitis pathogens. Results from 68 studies were included in the analysis and a lot of differences between studies were observed. The main conclusion was that observational studies did not show an effect (although they had the largest sample sizes), whereas challenge studies showed significant protective effects. The main problem of the challenge studies is the bypass of the natural defenses of the teat, the mastitis pathogens were often infused directly into the teat canal or teat cistern. Challenge studies that used teat immersion showed less protective effects than those who used infusion.

The often claimed protective effect of an IMI with CNS or *C. bovis* is not yet fully biologically explained. Sometimes the production of inhibitory substances is suggested (De Vliegher et al., 2004) and sometimes a competition for rate-limiting substrates in the milk is suggested as an explanation (Hogan et al., 1987). Also the moderate increase of the SCC after IMI with CNS or *C. bovis* is suggested as an explanation (Brooks and Barnum, 1984). Unfortunately, the IMI dataset did only contain monthly measures of SCC based on composite milk samples. There is no information on quarter level. An analysis with both SCC on quarter level and infection status with CNS or *C. bovis* as timevarying covariates in the same model is an interesting topic for further research but this is only possible if a similar dataset can be found with infection status and monthly SCC on quarter level. Such an analysis can give an idea of the relative importance of both factors and can

lead to new insights in the dynamics of IMI.

Another interesting topic to investigate is the impact of the width of the interval and the impact of one missing sample during summer (either July or August). In the dataset and the analysis we assumed that the infection can be detected until 1 month (or 2 months during summer) after infection. This is the case for persistent infections, but infections with some bacteria (e.g. enterobacteriaceae) are efficiently eliminated due to the immune response after a short period. The majority of the Gram-negative bacterial infections had a duration of less than 28 days (Todhunter et al., 1991). This means that some of the infections will be missed when the interval between sampling is 30 or in some cases even 60 days. A correction for this should be implemented. This also means that the results for *E. coli* and other coliforms are limited to the observed infections and that some of the infections will be missed.

Clustering of infections was high for all causal agents studied, this indicates large differences between the cows in their susceptibilty to new IMI. The clustering obtained in chapter 3 and 4 is due to differences between cows, but also between farms, unfortunately there is no quantification of the relative contribution of both. The stratified analysis on a subset of the data suggests that a big part of the clustering is due to differences between farms. In case of *S. aureus* the variance of the frailties decreased from 5.578 (cow+farm) to 2.699 (only cow effects) and for *S. uberis* from 5.852 to 3.989. Unfortunately, the stratified analysis was only possible on a subset of the data (only farms with more than 3 events), the conclusions are limited to those farms.

The relative importance of clustering within cow and within farm may suggest that a high clustering on cow level is an indication of a high within cow transmission, which is mainly the case for contagious pathogens. A high clustering on farm level may suggest that there are big differences between the farms. This means that those pathogens occur more frequent in some farms, a typical farm problem. This suggest that some other farms don't have problems with those pathogens. Additionally sanitary measures in problem farms can probably reduce the occurrence of those pathogens. Further research to identify risk factors on farm level is an interesting topic for further research.

5.2 Statistical methodology

The overall conclusion of sections 1.2.3, 1.3.5 and 1.3.6 is that data collected to study infection dynamics is in most cases longitudinal data. Although most studies have a time to infection nature, the original data is almost always reduced to binary data before analysis. This reduces the available information (time to infection) of the dataset but makes it possible to use commonly know analysis techniques (e.g. logistic regression, poisson regression) which are available in most commercial statistical software packages and where the interpretation of the estimated parameters is straight forward.

In this dissertation, the shared frailty model is introduced, taking into account more information than a univariate logistic regression or poisson regression. It models the evolution of the hazard of infection over time, uses the available information of censored observations and corrects for clustering between observations. This gives an idea of the evolution of the hazard of infection over time and provides an idea of the importance of clustering.

5.2.1 A piecewise constant hazard model with calendar time

The main objective of the analysis of the IBR-dataset was to model the evolution over calendar time. In most studies the start since entry in the study is modeled. In the IBR dataset, the start of the study is the implementation of a hyperimmunization protocol. A lot of animals enter the study at a time later than 0. The hazard function was integrated from the entrance time of the animal in the study population until the event or censoring time. This approach made it possible to model the hazard to seroconversion over calendar time in a study population with a natural turnover of cattle (birth of calves and purchase of cattle are balanced against death, culling and sale) which resulted in a continuous entrance of cattle into and exit of cattle from the study population.

The second problem was the fluctuating seasonal effect. The infection pressure seemed to be higher during winter periods (stabling period) than during summer periods (grazing period). This means that the baseline hazard will be higher during winter periods. This problem was solved by the implementation of a piecewise constant baseline hazard, allowing the baseline hazard to fluctuate seasonally.

The evolution of the difference over calendar time between the HIGs and the NIG was modeled, correcting for clustering and seasonal effects. The hyperimmunization resulted in a lower infection rate over time, a negative $\beta_{HIG.}$, the evolution of the hazard to seroconvert in the HIG compared to

the NIG was observed.

The proposed model is easy to implement and the results are straightforward. The results, however, were difficult to compare to other vaccination studies. Preliminary vaccination studies often only look to the antibody response and model the antibody titers compared to a control group. Also virus titers after artificial reactivation was sometimes measured. This is good in an experimental setting, but extrapolation to the field is difficult.

In most vaccination field studies the results are only descriptive. Most of them are based on observational data. A randomized study in different farms is a better choice and makes it possible to compare the new vaccine or protocol to a placebo treatment or the common used protocol. The hazard to seroconversion (which is the result of an infection) is modeled using a survival model and the HR can easily be obtained. Unfortunately, this type of modeling is not (yet) common in vaccination studies in veterinary science. In veterinary science the percentage survival/protected (after a pre-defined period) or the basic reproduction number (R_0) is used in most studies. A search on the web of science until august 2012 (based on the info in the title, abstract and keywords with search Topic=((vaccination or vaccine) and ((survival or frailty or cox) and model)) Refined by: Research Areas=(VETERINARY SCIENCES)) revealed that in the last 10 years only around 10 articles about vaccination were published where survival modeling was used (out of 69 results, most studies did not perform survival analysis, they only contained the word survival in another context). The use of survival models in human medicine is more common.

The percentage survival/protected after a pre-defined period is often used as outcome variable. The big problem with such analysis is that one has to determine the best or most meaningful time point, which is difficult. If multiple pre-defined endpoints are used, an appropriate correction for multiple comparisons should be performed. In most cases there are no corrections for drop-out of objects. If the drop-out is not balanced between the groups, it can also lead to bias.

The use of the basic reproduction number is also commonly used. The problem with the basic reproduction number is its lack in transparency. It is an artificial entity that can be estimated in several ways. Depending on the estimation method, the estimates can vary and comparing the R_0 from two different studies is inappropriate if the estimation procedure mentioned in the papers is not the same.

A nice topic for further research is a review of the experimental designs and statistical analyses used in the veterinary vaccine research. A lot of measures are used but most of them are less informative than the results

of a survival model. A re-analysis of the data of published studies can be performed with an appropriate survival model and the results can be compared with the originally published results. If the results of the new and old analysis are inconsistent, the reason for this inconsistency should be investigated (e.g. ignoring clustering, seasonal effects, drop-out etc. which can lead to bias)

5.2.2 Shared frailty model for interval-censored survival data with time varying covariates

In veterinary research, a lot of studies are interval censored due to feasibility reasons as daily sampling is sometimes expensive or impossible. In the IMI dataset, bacteriological sampling was performed each month. The problem of interval censoring can be solved in different ways. The most common one is by assuming the midpoint of the interval as an exact event time. Midpoint imputation leads to bias (slight overestimation of the fixed effect for parity and shape parameter, slight underestimation of the scale parameter of the Weibull distribution, sometimes underestimation of the standard errors of the parameter estimates). The shared gamma frailty model for right censoring was extended to handle interval censoring. It was still possible to integrate out the frailties in an analytical way and the marginal log likelihood can be maximized by general purpose maximizers. The inverse of the hessian matrix also provides the variance-covariance matrix of the estimated parameters.

To analyze the effect of an IMI with CNS (or *C. bovis*) another extension of the model was needed. The infection status with CNS is in other papers often assumed to be constant over time. In reality this is not the case, an IMI occurs at some time, the infection persists for some time and cure is also possible. This results in a time depending infection status. The conditional cumulative hazard was the sum of the cumulative hazard during the different intervals with a constant value for infection status. The extension to include time varying covariates was validated with a small simulation study and the results showed that it is an appropriate technique to analyze interval-censored clustered survival date with time varying covariates.

In the proposed model, some assumptions are made. The frailties are assumed to come from a one-parameter gamma distribution. Although Hsu et al. (2007) suggest that misspecification of the frailty distribution results generally only in low bias, a standard technique to test this assumption would be nice. Unfortunately, until now, no standard technique to diagnostics of the distribution of frailties has been generally accepted.

Another assumption is noninformative censoring. For most animals this will be the case. The end of lactation or the end of the study period is not related with the real time of infection. But what in case of culling due to an acute severe mastitis. Burvenich et al. (2007) state that nearly 25% of the cows with a severe gram-negative IMI in early lactation will either die or be culled. Some of those animals do not have a positive sample before culling because of the very acute nature if this clinical mastitis. In this case censoring is not independent from the event. Huang and Wolfe (2002) proposed a frailty model with informative censoring for right censored observations. This model, or an extension to interval-censoring, can be used to analyze the IMI dataset and the results can be compared to the original results to see if noninformative censoring is an acceptable assumption.

Although Radke (2003) demonstrated that ignoring interval censoring can lead to biased estimates, the relative importance of the bias due to interval censoring was modest when the results of the analysis with midpoint imputation are compared to the results of the model for interval censoring. Radke (2003) divided interval censoring in two types: synchronous (all intervals have comparable length and little or no overlap) and asynchronous (intervals of different cows may overlap, different lengths are observed: short intervals and long intervals). Bias due to asynchronous interval censoring is much higher than bias due to synchronous sampling. In the IMI dataset, the intervalcensoring is mainly synchronous (but also asynchronous: a longer interval during summer period, one missing sample during either July or August). This can be an explanation of the observed low bias due to interval censoring in the IMI dataset.

The current model only allows one frailty level. However, the data has two levels of clustering: the quarter is clustered in cow, the cow is clustered in farm. This can't be solved in an analytical way anymore. To have an idea of the clustering on cow and on farm level, a stratified model was fitted on a subset of the data in Chapter 3. This resulted in a frailty due to clustering within cow without common farm factors. Because a part of the variation will be due to farm factors, a decrease of the variance of the frailties is expected. In case of *S. aureus* the variance of the frailties decreased from 5.578 to 2.699 and for *S. uberis* from 5.852 to 3.989.

Ignoring the clustering on a higher level, farm level, has no meaningfull influence on the estimates of the fixed effects on a lower level, cow or quarter. But ignoring clustering on the level of the fixed effect can result in serious bias (Rondeau et al., 2006). There is also a second dataset with farm characteristics which can be merged with the IMI dataset. But for the analysis of risk factors on herd level, clustering on farm level has to

be included in the analysis to prevent bias due to the clustering on herd level. Unfortunately, no analytical solutions for this problem are available yet, other solutions have to be explored.

One possible solution for the two levels of clustering is the use of Bayesian statistics. Using a Gibbs-sampler, e.g. OpenBugs, WinBugs or Jags, a bayesian model with two levels of clustering can be fitted. This is an implementation of the model proposed by Wong et al. (2005) which can be easily adapted to the IMI dataset. However, the Bayesian approach is very computer intensive.

Another possible approximate solution can be obtained by integrating out the next level of the frailties in a numerical way. This is a possible extension of the current model which is not implemented yet for intervalcensored data. A similar approach is available for right-censored data with exact event times. It is available in the CRAN-package frailtypack (Rondeau et al., 2012). If interval censoring is ignored and the midpoint of the interval is assumed to be an exact event time, the data can be fitted using frailtypack. Unfortunately, the package didn't work on the full IMI dataset. The function in the package was able to analyze a small subset of the data after some time. But when the full dataset was analyzed, the computing started normal. Unfortunately, more than 3 months later, still no convergence was reported by the program, and we interrupted the process.

The implementation of the proposed model in R has still some problems. Due to the complex log likelihood, which contains a kronecker product, the maximum number of events allowed in one cluster, with the current code, is around 10 events (depending on the memory capacity of the computer). If there are more than 10 events in some clusters, when our model fails, the model can also be fitted in an approximate way using numerical integration. Bellamy et al. (2004) proposed a similar model for intervalcensoring in SAS with a log normally distributed frailty. This is a good alternative for the method proposed in this thesis.

The kronecker product issue should be solved in the future for application on trials with bigger cluster sizes. This needs a restyling of the function to calculate the kronecker product. The function should also be adapted to make use of the Surv() object in R and S-plus. The final full functional function (high number of events, time varying covariates) is then ready to be distributed as open source code by submission to CRAN (a network of R distributions and R packages and the contributed source code). In this way the function is available to all R-users.

Bibliography

Bellamy S.L., Li Y., Ryan L.M., Lipsitz S., Canner, M.J. and Wright R., 2004. Analysis of clustered and interval censored data from a community-based study in asthma. Stat. Med. 23, 3607-3621.

Brooks B.W. and Barnum D.A., 1984a. Experimental Colonization of the Bovine Teat Duct with *Corynebacterium bovis* and the Effect on Milk Somatic Cell Counts. Can. J. Comp. Med. 48, 141-145.

Burvenich C., Bannerman D., Lippolis J.D., Peelman L., Nonnecke B.J., Kehrli M.E. and Paape M.J., 2007. Cumulative physiological events influence the inflammatory response of the bovine udder to Escherichia coli infections during the transition period. J. Dairy Sci. 90, E39-E54.

De Vliegher S., Opsomer G., Vanrolleghem A., Devriese L.A., Sampimon O.C., Sol J., Barkema H.W., Haesebrouck F. and de Kruif A., 2004. In vitro growth inhibition of major mastitis pathogens by *Staphylococcus chromogenes* originating from teat apices of dairy heifers. Vet. Microbiol. 101, 215-221.

Hogan J.S., Pankey J.W. and Duthie A.H., 1987. Growth Responses of *Staphylococcus aureus* and *Streptococcus agalactiae* to *Corynebacterium bovis* Metabolites. J. Dairy Sci. 70, 1294-1301.

Hsu L., Gorfine M. and Malone K., 2007. On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is mis-specified. Stat. Med. 26, 4657-4678.

Huang X.L. and Wolfe R.A., 2002. A frailty model for informative censoring. Biometrics 58, 510-520.

Jaarverslag DGZ, 2009. Dierengezondheidszorg Vlaanderen, Drongen.

Jaarverslag DGZ, 2011. Dierengezondheidszorg Vlaanderen, Drongen.

Piessens V., De Vliegher S., Verbist B., Braem G., Van Nuffel A., De Vuyst L., Heyndrickx M. and Van Coillie E., 2012. Intra-species diversity and epidemiology varies among coagulase-negative Staphylococcus species causing bovine intramammary infections. Vet. Microbiol. 155, 62-71.

Radke B.R., 2003. A demonstration of interval-censored survival analysis. Prev. Vet. Med. 59, 241-256.

Reyher K.K., Haine D., Dohoo I.R. and Revie C.W., 2012. Examining the effect of intramammary infections with minor mastitis pathogens on the acquisition of new intramammary infections with major mastitis pathogensA systematic review and meta-analysis. J. Dairy Sci. 95, 6483-6502.

Rondeau V., Filleul L. and Joly P., 2006. Nested frailty models using maximum penelized likelihood estimation. Stat. Med. 25, 4036-4052.

Rondeau V., Mazroui Y. and Gonzalez J.R., 2012. Frailtypack: An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation. J. Stat. Softw. 47, 1-28.

Todhunter D.A., Smith K.L., Hogan J.S. and Schoenberger P.S., 1991. Gram-negative bacterial infections of the mammary gland in cows. Am. J. Vet. Res. 52, 184-188.

Wong M.C.M., Lam K.F. and Lo E.C.M., 2005. Bayesian analysis of clustered interval-censored data. J. Dent. Res. 84, 817-821.

Summary

In this thesis, the frailty model methodology is extended to address relevant questions in two important veterinary diseases: infectious bovine rhinotracheitis (IBR) and intramammary infections (IMI).

In Chapter 1 the datasets are introduced and a review of frailty model methodology is given. Survival analysis it typically used to analyze time-to-event data. A typical property of survival data is the occurrence of censoring. Right censoring occurs when no event was observed before the end of the follow up period (e.g. end of the study period or dropout due to another reason than the event). In this thesis also interval censoring is considered, hence the event time is contained in the interval between the last observed time without the event and the first time with the event, but the event time itself is unknown. The starting point of survival modeling in this thesis is the parametric shared frailty model, using Weibull distributed event times.

The frailty is introduced to deal with clustering. The 'shared' frailty model denotes the fact that the fraily is common for all subjects in the same cluster. An average cluster has a frailty of 1 while a frail (or strong) cluster has a frailty higher (or smaller) than 1. Frailties are assumed to follow a specific distribution. We will only make use of the one parameter gamma distribution for the frailty. This leads to a conditional (on the frailty) likelihood, but the frailty can be integrated out (when gamma distributed) and a marginal likelihood without the frailties can easily be obtained. The models used in this thesis are extensions of this parametric shared frailty model.

In Chapter 2 the effect of hyperimmunization on the transmission of infectious bovine rhinotracheitis (IBR) virus in cattle herds is investigated. The objective of the study is to assess the long-term effect and the effects of risk factors on the efficacy of two hyperimmunization protocols against IBR in a longitudinal field study during 28 months on 72 dairy and dairy-beef mixed farms with in total approximately 7700 cows.

The farms were assigned randomly to 3 treatment groups (hyperimmunization group (HIG) 1 and 2, which were hyperimmunized with glycoprotein E (gE)-deleted marker vaccines, and a nonintervention group (NIG)). Cattle in HIG 1 were initially vaccinated with an attenuated vaccine, whereas cattle in HIG 2 were initially vaccinated with an inactivated-virus vaccine. Cattle in both HIGs received booster inoculations with inactivated-virus vaccines at approximately 6-month intervals. The hazard for gE seroconversion was compared among experimental groups via a shared frailty model with a piecewise constant baseline hazard. The calendar time since the start of the hyperimmunization protocol was modeled, in order to see the evolution over time. The baseline was allowed to vary each season to correct for seasonal

and secular effects.

The hazard for gE seroconversion decreased significantly over time for the 2 HIGs, compared to the results for the NIG. Seasonal changes in the hazard of gE seroconversion were detected, with a higher risk in the winter periods than in the grazing periods. No significant difference was detected between HIG 1 and HIG 2. The only significant risk factor was the number of buildings for cattle on a farm; the higher the number of buildings, the lower the risk for gE seroconversion. The mean IBR prevalence decreased over time in both HIGs but remained constant or increased in the NIG.

Hyperimmunization via repeated administration of attenuated and inactivated virus gE-deleted marker vaccines as well as inactivated virus vaccines may provide a method for the control of IBR in cattle. Only in the bulls of the dairy-beef herds no significant decrease over time was observed which suggest that this will be the biggest challenge in the eradication process, but it is also the group with the highest turnover.

The proposed hyperimmunization protocols are used as common practice now, and a follow-up research of the hyperimmunization protocols is an interesting topic for further research.

In Chapters 3 and 4 the intramammary infections (IMI) dataset is analysed. Intramammary infections can result in (sub)clinical mastitis. Mastitis is the result of the inflammatory response mainly due to the invasion and replication of bacteria. IMI with some bacteria (*S. aureus* and *S. dysgalactiae*) result almost always in clinical mastitis while others (e.g. coagulase-negative Staphylococci (CNS) and *Corynebacterium bovis*) only induce a mild inflammation response, an often unobserved subclinical mastitis.

In a longitudinal observational study with in total 1132 cows on 25 dairy herds, the time to infection with mastitis pathogens was measured based on monthly milk samples at quarter level. This results in interval-censored time to event data. The infection status with CNS and *C. bovis*, a time varying covariate, was also registered.

Udder infections in dairy cows are observed at udder quarter level. Therefore, the best strategy to study infection dynamics of particular bacteria causing mastitis is to follow up and model individual udder quarter infection times. Udder quarter infection times, however, are not independent as they are clustered within a cow and herds. Another challenge in modeling the infection times is the interval censoring; it is only known that the infection has taken place in the interval between the last negative and the first positive sample.

We applied a technique based on the gamma frailty model which handles

the clustering and interval censoring simultaneously. Parameter estimates can be obtained analytically and their variance is obtained by the inverse of the hessian matrix. The proposed technique was applied to udder quarter infection times for *C. bovis*, *Staphylococcus aureus* and *Streptococcus uberis*. Multiparous cows were more likely to get infected earlier in lactation with *C. bovis* or *S. uberis* than primiparous cows. The times to infection of all three bacteria were highly clustered at cow level and the results of a stratified model on a subset of herds suggested a high clustering on herd level for *C. bovis* and *S. uberis*.

In Chapter 4, the influence of intramammary infections (IMI) with CNS or *Corynebacterium bovis* on the hazard of infection with other mastitis pathogens was investigated. Frailty model methodology was used to analyze the time to event data taking into account clustering (quarters in udders), interval censoring and the time varying CNS or *C. bovis* infection status. The presented model was validated in a separate simulation study based on 500 datasets and suggested that the proposed technique was appropriate to analyze the data.

Quarters infected with CNS were more vulnerable to an IMI with *S. uberis* (HR: 2.02, $P = 0.005$) and *S. dysgalactiae* (HR: 3.29, $P < 0.001$) than uninfected quarters. On the other hand, an IMI with *C. bovis* did have a significant protective effect against an IMI with *S. aureus* (HR: 0.54, $P = 0.002$), *S. uberis* (HR: 0.22, $P < 0.001$) or *S. dysgalactiae* (HR: 0.32, $P = 0.004$). No significant effect of an IMI with *C. bovis* on the occurrence of *E. coli* and other coliform bacteria was found ($P > 0.37$).

Samenvatting

In deze thesis is de frailty model methodologie uitgebreid om analyzeproblemen op te lossen bij twee relevante diergeneeskundige aandoeningen: infectieuze bovine rhinotrachetis (IBR) en intramammaire infecties (IMI).

De datasets en de basis voor survival analyse zijn beschreven in Hoofdstuk 1. Survival analyse is typerend om de tijd tot een bepaalde gebeurtenis te modelleren en censurering is hierbij kenmerkend. Rechter censurering komt voor als er geen gebeurtenis is waargenomen op het einde van de opvolgtijd (bv. einde van de studietijd, het verlaten van de studie door een andere reden dan de gebeurtenis). Wanneer de exacte tijd van de gebeurtenis niet gekend is, dan spreken we over interval censurering. Hierbij ligt de exacte tijd tussen de tijd van de laatste negatieve test en de eerste positieve test. In deze thesis is vertrokken van het parametrisch shared frailty model waarbij Weibull verdeelde tijden worden verondersteld.

Als de overlevingstijden van de verschillende subjecten niet onafhankelijk zijn van elkaar, kan een frailty worden gebruikt om de clustering te modelleren. In het shared frailty model is er een gemeenschappelijke frailty term voor alle subjecten in dezelfde cluster. Een gemiddelde cluster heeft dan een frailty met waarde 1 terwijl een zwakkere (of sterkere) cluster een frailty heeft die groter (of kleiner) is dan 1. De frailties worden verondersteld een distributie te volgen. In deze thesis wordt enkel met de één parameter gamma distributie gewerkt. Dit leidt tot een voorwaardelijke aannemelijkheidfunctie (gekende frailty) waarvan de frailty kan worden uitgeïntegreerd wat resulteert in een marginale aannemelijkheidfunctie. De modellen die in deze thesis worden gebruikt, zijn uitbreidingen van het parametrisch shared frailty model voor rechts gecensureerde tijden.

In Hoofdstuk 2 wordt het langetermijn effect van hyperimmunisatie met een marker vaccin tegen infectieuze rhinotracheitis onderzocht. Het doel van de studie is om zowel de lange termijn effecten, als risico factoren op de effectiviteit van twee hyperimmunisatie protocollen na te gaan. Hiervoor is een grote longitudinale veldstudie van 28 maand op 72 bedrijven uitgevoerd, met bloedafnames bij meer dan 7700 koeien.

Aan de bedrijven is willekeurig 1 van de drie behandelingen toegekend. Hyperimmunisatie groep 1 (HIG 1) is initieel met een geattenueerd vaccin gevaccineerd terwijl hyperimmunisatie groep 2 (HIG 2) initieel met een geïnactiveerd vaccin werd gevaccineerd. Beide hyperimmunisatie groepen hebben om de 6 maanden booster vaccinaties met een geïnactiveerd vaccin toegediend gekregen. Bij de derde groep hebben de bedrijven hun gebruikelijk vaccinatieschema gevolgd (non intervention group, NIG). Alle vaccins zijn glycoprotein E (gE)-deletie marker vaccins. De ratio van het risico op gE seroconversie van de hyperimmunisatie groepen ten opzichte van de

NIG is gemodelleerd met een shared frailty model met een constant basisrisico dat kan variëren tussen de verschillende seizoenen zodat seizoens- en jaarverschillen meegenomen zijn in de analyse.

Het risico op gE seroconversie is in vergelijking met de NIG significant gedaald tijdens de studietijd in de twee hyperimmunisatiegroepen. Er zijn seizoensvariaties waarbij het risico op seroconversie groter is tijdens de winterperiode ten opzichte van de zomerperiode (de weideperiode). Er is echter geen significant verschil vastgesteld tussen de twee hyperimmunisatiegroepen. De enige significante risicofactor blijkt het aantal rundveestallen op een bedrijf: hoe meer stallen, hoe lager kans op seroconversie. De gemiddelde seroprevalentie is tijdens de studie in de twee hyperimmunisatiegroepen gedaald, maar blijft constant of stijgt zelfs lichtjes in de NIG.

Hyperimmunisatie met herhaalde vaccinaties met geattenueerde vaccins gevolgd door geïnactiveerde gE-deletie marker vaccins, alsook met enkel geïnactiveerde gE-deletie marker vaccins is dus een mogelijke controlemaatregel in de strijd tegen IBR. Enkel bij de stierenpopulatie van gemengde bedrijven (melkvee-vleesvee) is er geen significante daling te merken. Dit is enerzijds een uitdaging in het eradicatieproces, maar anderzijds is het ook de groep met de hoogste turnover.

De voorgestelde hyperimmunisatie protocols zijn ondertussen algemeen gebruikt in de praktijk en een opvolgstudie zou een interessante topic zijn voor verder onderzoek.

In de Hoofdstukken 3 en 4 is de intramammaire infecties (IMI) dataset geanalyseerd. Mastitis is het gevolg van een ontstekingsreactie ten gevolge van een invasie en vermenigvuldiging van bacteriën. Een IMI met sommige bacteriën (zoals *S. aureus* en *S. dysgalactiae*) resulteert bijna altijd in een klinische mastitis terwijl andere (zoals CNS en *Corynebacterium bovis*) vaak een onopgemerkte subklinische mastitis tot gevolg hebben met een heel milde ontstekingsreactie.

In een longitudinale observationele studie op 25 melkveebedrijven, met in totaal 1132 melkkoeien, is de tijd tot infectie met mastitis pathogenen geschat aan de hand van maandelijkse melkstalen per kwartier. Dit resulteert in intervalgecensureerde infectietijden. De infectiestatus met CNS en *C. bovis*, een tijdsafhankelijke variabele, is eveneens geregistreerd.

Intramammaire infecties zijn per kwartier geregistreerd zodat het modelleren van de infectie dynamiek van een bepaalde bacterie per kwartier de beste keuze is. Uierkwartier infectietijden zijn echter niet onafhankelijk van elkaar, ze zijn geclusterd binnen een koe, en de koeien zijn geclusterd binnen bedrijven. Een ander probleem is dat de infectietijden niet exact gekend zijn, maar interval gecensureerd. We weten alleen dat er een infectie heeft

plaatsgevonden tussen het laatste negatieve staal en het eerste positieve.

Voor deze analyse is een shared gamma frailty model gebruikt dat rekening houdt met de clustering van de kwartieren binnen een koe en met de intervalcensoring van de infectietijden. Schatters voor de parameters kunnen op een analytische manier worden bekomen en hun variantie kan worden berekend door de inverse van de hessian matrix te nemen. De voorgestelde techniek is toegepast op de uierkwartier infectietijden met *C bovis*, *Staphylococcus aureus* en *Streptococcus uberis*. Multipare koeien blijken sneller geïnfecteerd te worden met *C. bovis* of *S. uberis* dan primipare koeien. Er is een grote clustering waargenomen voor alle drie de bacteriën op koe niveau en de resultaten van een gestratificeerd model, gefit met een aantal van de bedrijven, suggereren een hoge clustering op bedrijfsniveau voor *C. bovis* en *S. uberis*.

In Hoofdstuk 4 wordt het effect van een IMI met CNS of *Corynebacterium bovis* op het risico op infectie met andere mastitis pathogenen onderzocht. De frailty model methodologie is gebruikt om de interval gecensureerde infectietijden te analyseren, waarbij rekening is gehouden met de clustering van de kwartieren binnen de koe en een tijdsafhankelijke variabele, namelijk de infectiestatus met CNS of *C. bovis*. Het model uit Hoofdstuk 3 is voor deze analyse van tijdsafhankelijke variabelen uitgebreid. Het model is gevalideerd met een afzonderlijke simulatiestudie gebaseerd op 500 gegenereerde datasets. Deze toont aan dat de techniek geschikt is om de data te analyseren.

Kwartieren die geïnfecteerd zijn met CNS blijken meer gevoelig voor een IMI met *S. uberis* (HR: 2.02, $P = 0.005$) en *S. dysgalactiae* (HR: 3.29, $P < 0.001$) dan niet geïnfecteerde kwartieren. Dit in tegenstelling tot kwartieren die geïnfecteerd zijn met *C. bovis* die minder gevoelig zijn voor een IMI met *S. aureus* (HR: 0.54, $P = 0.002$), *S. uberis* (HR: 0.22, $P < 0.001$) of *S. dysgalactiae* (HR: 0.32, $P = 0.004$). Er is geen significant effect vastgesteld van een IMI met *C. bovis* op het voorkomen van een IMI met *E. coli* of andere coliforme bacteriën.

Dankwoord

Een doctoraat is het resultaat van jaren werk. Tijdens die jaren kruisten veel mensen mijn pad, zo veel dat het zelfs gevaarlijk wordt ze op te noemen (omdat ik er zeker vergeet), maar toch doe ik een poging!

Vooreerst wil ik mijn promotor Prof. dr. Luc Duchateau danken. Je hebt me de kans gegeven me verder bij te scholen in de statistiek. Onder jouw begeleiding kon ik dit doctoraat tot een goed einde brengen. Jouw expertise over frailty models, ondertussen in boekvorm gegoten, en eindeloze ideeën hebben de aanzet gegeven voor dit doctoraat. Ook onze geanimeerde discussies over allerhande levensbeschouwelijke onderwerpen zullen me nog lang bijblijven.

Daarnaast wil ik ook mijn bureaugenote van het eerste uur, Dr. Klaartje Goethals, danken. Voor alle wiskundige steun en deskundige toelichting over integralen e.d.. Eveneens bracht je me tijdens de opleiding biostatistiek aan de UHasselt de inzichten bij van de cursus statistical inference, wat zonder jouw hulp niet mogelijk was geweest. We hebben de afgelopen jaren veel mooie, maar soms ook mindere momenten beleefd. Binnenkort zullen we dan wel geen collega's meer zijn, maar ik hoop van harte dat we vrienden blijven.

Ik wil ook een speciaal woord van dank richten aan alle co-auteurs van de artikels in dit proefschrift. In volgorde van verschijnen zijn dit: Prof. dr. Niko Speybroeck , Prof. dr. Dirk Berkvens, Alain Dupont, dr. Pierre Kerkhofs, Prof. dr. Etienne Thiry , dr. Marc Dispas en dr. Hans Laevens.

Onder hen wil ik een paar mensen extra danken. Dr. Marc Dispas, wil ik danken voor het ter beschikking stellen van de IBR dataset voor verder onderzoek. Hartelijk dank ook voor de vele opbouwende commentaren en de vele suggesties tot op het einde.

Dr. Hans Laevens wil ik danken voor het ter beschikking stellen van de mastitis dataset en voor de interessante feedback op de artikels over mastitis. Ik vond het altijd heel boeiend om met je te brainstormen over de mogelijkheden van de dataset en het biologisch interpreteren van de vele resultaten.

Prof. dr. Christian Burvenich, bedankt voor de vele opmerkingen en suggesties over mastitis. Je bent een vat vol creatieve ideeën en hebt de gave mensen blijvend te motiveren, ook als het even wat moeilijk wordt. Je feedback begon altijd met een paar vrolijke noten waardoor aanpassingen niet meer onoverkomelijk leken.

Graag ook een woord van dank aan de lees- en examencommissie: Prof. dr. Frank Gasthuys, Prof. dr. Christian Burvenich, Prof. dr. Sarne De Vliegher, dr. Marc Dispas, Prof. dr. Dominiek Maes, Prof. dr. Ziv Shkedy and Prof. dr. Stijn Vansteelandt. Jullie suggesties brachten me tot inzichten

die hebben bijgedragen aan de finale versie van dit proefschrift.

Bovendien wens ik ook al mijn collega's en ex-collega's te danken voor de aangename werksfeer in de vakgroep.

Ik wil graag mijn ouders danken. Toen ik begon aan mijn opleiding Diergeneeskunde sprak ik lachend over een tienjarenplan. Ziehier, het zijn er veel meer geworden... maar nu kan je eindelijk zeggen dat jullie jongste zoon is afgestudeerd!

Tot slot wil ik ook mijn trouwe steun en toeverlaat danken, Karen. Je hebt me altijd gesteund tijdens de momenten dat het wat moeilijker was en leren relativeren wanneer nodig. Je wist natuurlijk als geen ander wat doctoreren is! Ik wil je ook danken om ons familieleven al die jaren in goede banen te leiden en om ons drie schatjes van kinderen te schenken. Elias, Theresa en Icarus wil ik danken omdat ze hun papa 'ongestoord' hebben laten werken, en ik beloof dat ik nu werk zal maken van jullie kamers.