SEMIPARAMETRIC HIERARCHICAL PROPORTIONAL HAZARDS MODELS WITH APPLICATIONS TO ANIMAL HEALTH DATA

BY

ADEL I. ELGHAFGHUF

A Thesis
Submitted to the Graduate Faculty
In Partial Fulfillment of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

Department of Health Management
Faculty of Veterinary Medicine
University of Prince Edward Island

CONDITIONS FOR THE USE OF THE THESIS

PERMISSION TO USE GRADUATE THESIS

Title of Thesis:

SEMIPARAMETRIC HIERARCHICAL PROPORTIONAL HAZARDS MODELS WITH APPLICATIONS TO ANIMAL HEALTH DATA

Name of Author: ADEL I. ELGHAFGHUF

Department: Health Management

Degree: Doctor of Philosophy Year: 2014

Name of Supervisor: Dr. Henrik Stryhn

In presenting this thesis in partial fulfilment of the requirements for a graduate degree from the University of Prince Edward Island, the author has agreed that the Robertson Library, University of Prince Edward Island, may make this thesis freely available for inspection and gives permission to add an electronic version of the thesis to the Digital Repository at the University of Prince Edward Island. Moreover the author further agrees that permission for extensive copying of this thesis for scholarly purposes may be granted by the professor or professors who supervised the author's thesis work, or, in their absence, by the Chair of the Department or the Dean of the Faculty in which the author's thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without the author's written permission. It is also understood that due recognition shall be given to the author and to the University of Prince Edward Island in any scholarly use which may be made of any material in the author's thesis.

Signature: _____

Address: 550 University Avenue, Charlottetown, PEI, C1A 4P3

Date: _____

University of Prince Edward Island

Faculty of Veterinary Medicine

Charlottetown

CERTIFICATION OF THESIS WORK

We, the undersigned, certify that **Adel I. Elghafghuf**, a candidate for the degree of Doctor of Philosophy has presented his thesis with the following title:

SEMIPARAMETRIC HIERARCHICAL PROPORTIONAL HAZARDS MODELS WITH APPLICATIONS TO ANIMAL HEALTH DATA,

and that the thesis is acceptable in form and content, and that a satisfactory knowledge of the field covered by the thesis was demonstrated by the candidate through an oral examination held on

Examiners:

_____ Dr.

_____ Dr.

_____ Dr.

_____ Dr.

_____ Dr.

Date: _____

ABSTRACT

This thesis discusses and applies hierarchical models for survival data in the field of veterinary medicine. The focus is on hierarchical proportional hazards models when the baseline hazard is left completely unspecified. Parameter estimation for these models is explored and the performance of their estimation methods is investigated in terms of statistical properties such as unbiasedness, robustness, and probability coverage.

The thesis is formed by manuscripts of four studies. The first study compares, via simulation, the performance of different estimation methods for estimating a random slope Cox model with and without covariance between the random effects. The simulation is built to mimic real animal health data. The aim of the study is to establish some practical guidelines for the choice of appropriate statistical estimation methods for modeling random slopes in 2-level hierarchical data. Results show that estimating the full covariance matrix for random effects is always preferable in the analysis and Poisson maximum likelihood estimation is an adequate approach for this task.

The second study explores the feasibility of a full hierarchical survival analysis for a large dataset with three levels of hierarchy and time-dependent predictors and coefficients. To this end, a log-normal nested frailty Cox model is applied to Canadian Bovine Mastitis Research Network (CBMRN) data to identify risk factors associated with the hazard of clinical mastitis (CM) during cow lactations. This nested frailty model is estimated by the Poisson maximum likelihood approach with Gaussian quadrature. The performance, in terms of bias and efficiency of estimates, of the Poisson maximum likelihood approach (estimated using either Gaussian quadrature or Laplace

approximation) is compared with the performance of the penalized partial likelihood approach. The Poisson maximum likelihood with Gaussian quadrature produces fairly robust and adequate estimates while the penalized partial likelihood and the Poisson maximum likelihood with Laplacian approximation are found to have substantial drawbacks. Further, the research indicates that some of the herd managerial factors combined with cow characteristics influence the hazard of CM during the lactation period; some of these effects are different earlier as compared to later in the lactation.

The third study involves analyzing a dataset on calf loss and mortality in beef cattle in Western Canada. This dataset has a cross-classified and multiple membership structure which is a special type of data structure that has only been accounted for in the analyses of linear and generalized linear models but not in survival analysis. The study objectives are twofold: the first is to explore and demonstrate the use of Poisson generalized linear mixed models (GLMMs) in the Bayesian framework for estimating a Cox model with cross-classified and multiple membership frailties. The second, is to simultaneously examine the individual, herd management, and environmental factors associated with beef calf mortality in Western Canada and to estimate the age period where calves are most at risk. Finally, a simulation study with settings similar to the real data is carried out to evaluate the estimation approach. The simulation results gave evidence that the approach used provides valid estimates.

In the fourth study, the robustness of Poisson maximum likelihood estimation was assessed, through simulation, for a Cox model with normal random effects under misspecification of the random-effects distribution. The impact of misspecifying the distribution of random effects is assessed based on two different non-normal distributions

for random effects and three different model designs. Some of the factors that might affect the estimation are also investigated. The study shows that the Poisson maximum likelihood approach yields robust estimates under misspecification of the random-effects distribution for within-group fixed effects and in a wide range of situations for between-group fixed effects. For variance components, the approach produces robust estimation under model misspecification as long as the magnitude of heterogeneity is small, though misspecification may become a matter of concern when the magnitude of heterogeneity and group sizes become large.

ACKNOWLEDGEMENTS

This is always a difficult part because many people have helped in many different ways. However, there are those that should be recognized for the critical help they provided. First and foremost, I would like to express my deepest gratitude to my supervisor Dr. Henrik Stryhn. From the first interview with him to the last revision of my thesis, from weekly meetings to individual discussions, I can always feel his enthusiasm and patience. Without his excellent guidance, constant encouragement, and endless support, this thesis would not have been possible.

I thank the Libyan Ministry of Higher Education and Scientific Research (MOHESR) and Canadian Bureau for International Education (CBIE) for my personal support and funding this research project.

I want to thank the Atlantic Computational Excellence Network (ACEnet) for access to computational facilities. My thanks are also extended to the Epi group at AVC for providing access to the Epi-informatics computer lab.

I would like to thank my supervisory committee members, Dr. Greg Keefe, Dr. Crawford Revie, Dr. Sami Khedhiri, Dr. Spencer Greenwood, Dr. Ian Dohoo, for their careful reading of this thesis, the good comments, and the helpful suggestions to improve this work.

Special thanks are due to Dr. Ian Dohoo, for sharing his experience and knowledge. His critical comments and advices have contributed so much for this thesis. I am also thankful to Dr. Spencer Greenwood and Dr. Sami Khedhiri for their advice and critical discussions of my thesis.

I would like to acknowledge Dr. Cheryl Waldner, Dr. Kristen Reyher and Dr. Simon Dufour for their fruitful collaboration. Thanks are also extended to Dr. Jette Christensen.

Thanks all colleagues, fellow graduate students, and staff in Centre for Veterinary Epidemiological Research and AVC for the excellent work environment.

Many thanks are due to the wonderful friends that I met them here, in Charlottetown. Thank you all.

I express my sincere thanks to my parents, brothers and sisters for their love, kindness, encouragement, patience and support.

Finally, special thanks go to my wife, Salma, who has been my true support and friend in this long journey of graduate studies. This thesis could not have been made without her constant support and encouragement. My son, Ibrahim, who’s sweet smile is an amazing power for me to complete this work.

Dedicated to:

Family and parents

CONTENTS

List of Tables

List of Figures

List of Abbreviations

ACEnet	Atlantic Computational Excellence network
AFT	Accelerated Failure Time
BAY	Bayesian approach
BLUP	Best Linear Unbiased Predictor
CBMRN	Canadian Bovine Mastitis Research Network
CI	Confidence Interval
CM	Clinical Mastitis
CMM	Cross-classified and Multiple Membership
DCT	Dry-off Cow Treatment
DHI	Dairy Herd Improvement
DIC	Deviance Information Criteria
DIM	Days In Milk
EM	Expectation Maximization
Esd	Empirical standard deviation
ETS	use external teat sealant at dry-off
GLLAMM	Stata program for estimating Generalized Linear Latent and Mixed Models
GLM	Generalized Linear Model
GLMM	Generalized Linear Mixed Model
HEPS	Health and Production Surveillance
HR	Hazard Ratio
IMI	Intra-mammary Infection
ITS	use internal teat sealant at dry-off
IQR	Interquartile range
MCEM	Monte Carlo with EM

MCMC	Markov Chain Monte Carlo
ML	Maximum Likelihood
MMA	Mastitis/Metritis/Agalactia
MSE	Mean Squared Error
NCDF	National Cohort of Dairy Farms
PC	Personal Computer
PFL	Penalized Full Likelihood
PH	Proportional Hazards
PML	Poisson Maximum Likelihood
PMLAP	Poisson Maximum likelihood with Laplace approximation
PMLGQ	Poisson Maximum likelihood with Gaussian quadrature
PPL	Penalized Partial Likelihood
REML	Residual Maximum Likelihood
SCC	Somatic Cell Count
SCS	Somatic Cell Score
SD	standard deviation
SE	Standard Error

Chapter 1

Introduction

1.1. Survival data

When data include a response variable that corresponds to the time from a well-defined origin to the occurrence of a particular event or end point, the data may be called time-to-event data. Such data have the following features: the response variable is non-negative; part of the data is often right censored, i.e., for some individuals the event of interest has not occurred due to withdrawal or end of study (there are also other types of censoring such as left and interval censoring); and the data may contain predictors and effects that change over time. Time-to-event data go under different names depending on the area of application, for example as survival data in medical and biological sciences and as failure time data in engineering and industry. Examples of time-to-event outcomes are: time from onset of a disease to death, time to recurrence of a disease, time to breakdown of a machine, or the duration time of a process.

In veterinary and animal sciences, survival data are often encountered. Knowing when the event of interest is going to occur and what factors might affect the instantaneous rate of this event may be of crucial importance for both veterinarians and farmers to take the necessary action. In the last several years, an increasing number of both experimental and observational studies using time-to-event response in veterinary and animal science have been observed for different event times including, for instance, time to death in different

species (Nur et al., 2004; Duncan et al., 2006; Burnley et al., 2010; Hatcher et al., 2010; McCorquodale et al., 2013), time to culling in cattle (Gröhn et al., 2005; Duchateau and Janssen, 2008; Dohoo et al., 2009), time from calving to conception in dairy cows (Meadows et al., 2006; Meadows et al., 2007), time to occurrence of a disease (Portolano et al., 2007; Nielsen and Dohoo, 2012; Relun et al., 2013), and time to recovery from a disease (Edmondson et al., 1989; Nielsen and Dohoo, 2013). The response variable also could be time to other events such as time to first application of treatment (Christensen, 1996) or concentration of antibiotic to event (time to event) where inhibition of bacterial growth is the event (Stegeman et al., 2006). In veterinary survival data, censoring may happen for various reasons. For instance, in studies of occurrence time of mastitis in dairy cows censoring could have occurred to the animal due to culling, drying off, leaving the study, or surviving until the end of the study without a clinical case occurring.

Survival analysis refers to statistical techniques and methods for modeling and analyzing time-to-event data. These techniques play an important role in many fields such as epidemiology, demography, medicine, engineering, and economics. The response variable in survival analysis is often modeled indirectly through the instantaneous event rate (hazard function) that can be linked to the survival function. The proportional hazards model (Cox, 1972) has become the preferred regression analysis for survival data. The Cox proportional hazards model is the main focus throughout this thesis. In the next section, a detailed description of three survival datasets from veterinary medicine is given.

1.2. Description of datasets used in the thesis

1.2.1. Lameness data

The lameness data are described in Christensen (1996) and in Josiassen and Christensen (1999). Briefly, a total of 7632 litters of piglets (5,465 sows) from 35 herds were followed from birth to weaning in a Danish project carried out by the Health and Production Surveillance System (HEPS) during the period from October 1990 to March 1991. All clinical signs related to the lameness such as splayleg, joint infection, or ataxia, were monitored in the litters and necessary treatment was applied and recorded by the producers. The event time for a litter was defined as the time from birth to the first application of treatment for lameness in the litter, while the censoring time was the time from birth until either the litter was weaned or excluded at 40 days of follow-up. In Chapter 2, data from the 22 herds are used as an example. These herds did not participate in any elevated health programs (such as specific pathogen free herds that are declared free from certain infectious diseases), and only the first litter per sow was included. The failure rate in these 22 herds was 11.2%. The predictor of primary interest here was sow treatment for milk fever, infection, or MMA (mastitis/metritis/agalactia) in days around farrowing (2 days before and up to 4 days after); 26% of the sows were treated.

1.2.2. CBMRN data

The Canadian Bovine Mastitis Research Network (CBMRN) data were from the National Cohort of Dairy Farms (NCDF) collected from January 2007 to December 2008. The herd selection process and data collection of NCDF are described in Reyher et al. (2011). In brief, a total of 8,035 cows from 69 herds were followed for 10,831 lactations until the cow experienced the first case of clinical mastitis (CM), was culled or dried off, left the

cohort, or its follow-up was interrupted by the end of the study. The response variable was defined as the number of days from calving until developing a first case of CM. Observations from cows that did not develop a mastitis case within the lactation were considered right censored. Important information at both the individual and herd levels, such as time of CM occurrence, calving date, culling date, dry-off date, lactation number, herd somatic cell score (SCS), and herd demographics were captured. Other information related to the herd management was provided by an udder health related management survey described in Dufour et al. (2010). The incidence of clinical mastitis in this dataset was 14.2%. The CBMRN dataset is analyzed in Chapter 3.

1.2.3. Calf mortality data

The calf mortality dataset was from the Western Canada beef productivity study (Waldner, 2008). The study was on calf loss and mortality in beef cattle in Western Canada where 23,409 calves from 174 herds were followed for up to 180 days after calving during the period from January-June 2002. The dataset included 897 cases of mortality, corresponding to an incidence of 3.8%. The event was defined as a case of calf mortality that occurred at least one hour after birth; the event time was defined as the time from calving to death (recorded in days). Calves that were sold during the follow-up period or survived until the end of the follow-up period were considered right censored observations. The calf mortality dataset is studied in Chapter 4.

1.3. Survival analysis fundamentals

Survival analysis is the analysis of data that correspond to the time from a well-defined time origin until the occurrence of some particular event or end-point. Let T_j and

C_j $(j = 1, \ldots, N)$ be independent variables representing the event and right censoring times for the j^{th} subject, respectively. Let $Y_j = \min(T_j, C_j)$ denote the observed time for the subject j and $\delta_j = I(T_j \leq C_j)$ be the event indicator taking 1 if the event is observed and 0 otherwise. Suppose that $f(t)$ is the probability density function of the event time T; then the survival function that describes the survival data is given by

$$S(t) = P(T > t) = 1 - \int_0^t f(s)ds = 1 - F(t) \tag{1.1}$$

where $F(t)$ is the cumulative distribution function. The survival function captures the probability that the subject will survive beyond a specified time.

We are interested in calculating the probability that a subject survives in the interval $[t, t + \Delta t)$ where $\Delta t > 0$, conditional on having survived to the beginning of that interval (to time t), so

$$P(t \leq T < t + \Delta t | T \geq t) = \frac{P(t \leq T < t + \Delta t)}{P(T > t)} = \frac{S(t) - S(t + \Delta t)}{S(t)} \tag{1.2}$$

Dividing (1.2) by Δt to obtain,

$$\frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{-1}{S(t)} \times \frac{S(t) - S(t + \Delta t)}{\Delta t} \tag{1.3}$$

Taking the limit of both sides of (1.3) as $\Delta t \to 0$, the hazard function is obtained:

$$\lambda(t) = \frac{-d[\ln S(t)]}{dt} = \frac{f(t)}{S(t)} \tag{1.4}$$

This hazard function can be interpreted as the instantaneous failure rate for a subject surviving to time t. From (1.4), $S(t)$ can be written as

$$S(t) = \exp(-\Lambda(t)) \tag{1.5}$$

where $\Lambda(t) = -\int_0^t \lambda(s)ds$ is the cumulative hazard function. This leads to $\Lambda(t) = -\ln S(t)$, and so $\hat{\Lambda}(t) = -\ln \hat{S}(t)$.

Let $t_j,\ j = 1, \dots, r$ denote the ordered, observed and distinct failure times, i.e. $t_1 < \cdots < t_r$. The Kaplan-Meier estimator (Kaplan and Meier, 1958) for the survival function $S(t)$ can be calculated as

$$\hat{S}(t) = \prod_{j:t_j \le t} \left(1 - \frac{d_j}{n_j}\right) \tag{1.6}$$

where the d_j denotes the number of failures at time t_j, and n_j is the number of subjects at risk at time t_j, i.e. the number of subjects still alive just before t_j. The Kaplan-Meier estimator is a step function, in which the estimated survival probabilities are constant between event times and decrease at each observed event time.

1.4. Regression models for survival data

Several approaches have been used in survival analysis to model the effect of explanatory variables on the time to event. Some of these models are presented briefly in the next sections. Let $\boldsymbol{x_j} = (x_{1j}, \dots, x_{pj})$ be a vector of explanatory variables for the j^{th} subject throughout.

1.4.1. Proportional hazards (PH) models

The most popular procedure to associate the hazard function $\lambda(t)$ and $\boldsymbol{x_j}$ is based on the concept of a proportional hazards (PH) model (Cox, 1972). In PH models, the hazard function is a product of a baseline hazard $\lambda_0(t)$ (where all the variables included in the model are zero) and a non-negative function of the explanatory variables $\varphi(\boldsymbol{\beta}'\boldsymbol{x_j})$. The most common and convenient choice for the non-negative function is $\varphi(\boldsymbol{\beta}'\boldsymbol{x_j}) = \exp(\boldsymbol{\beta}'\boldsymbol{x_j})$. The PH model can then be written as

$$\lambda_j(t) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x_j}) \tag{1.7}$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients associated with $\boldsymbol{x_j}$.

The proportional hazards terminology means that the hazard ratio (HR) of two subjects, say I and II with explanatory variables x_I and x_{II}, does not depend on time, i.e.,

$$HR = \frac{\lambda_I(t)}{\lambda_{II}(t)} = \frac{\lambda_0(t)\exp(\beta x_I)}{\lambda_0(t)\exp(\beta x_{II})} = \exp[\beta(x_I - x_{II})],$$

is constant over time.

The baseline hazard $\lambda_0(t)$ of model (1.7) can be modeled parametrically and the most common approach is a Weibull PH model, or nonparametrically by using a PH Cox model (Cox, 1972). The latter is probably the more widely used one. In the Cox PH model, no assumptions about the form of $\lambda_0(t)$ are made (non-parametric part of the model) but a parametric form for the effect of the predictors (parametric part of the model) is needed. Therefore, the model is referred to also as a semi-parametric model. The attractive feature of the semi-parametric approach is that an adequate estimate for the

regression coefficients $\boldsymbol{\beta}$ can be obtained even though the baseline hazard is not specified.

In the PH Cox model, inference on $\boldsymbol{\beta}$ can be made by partial likelihood (Cox, 1972), which is a part of the full likelihood that does not depend on $\lambda_0(t)$. Suppose we observe $(y_j, \delta_j, \boldsymbol{x_j})$ for subject j, where y_j is an observed time, δ_j is an event indicator, and $\boldsymbol{x_j}$ is a vector of explanatory variables. As described in Collett (1994), the probability that the j^{th} subject fails at some time t_j conditional on t_j being one of the observed set of r failure times $t_1, t_2, \dots, t_r$ is

$$\Pr(\text{subject } j \text{ with } \boldsymbol{x_j} \text{ fails at } t_j \,|\, \text{one failure at } t_j) = \frac{\Pr(\text{subject } j \text{ with } \boldsymbol{x_j} \text{ fails at } t_j)}{\Pr(\text{one failure at } t_j)}$$

The numerator of the above equation is the hazard of event at time t_j for subject j with explanatory variables $\boldsymbol{x_j}$, this hazard function can be written as $\lambda_j(t_j)$. The denominator is the sum of the hazards of event at time t_j. This is the sum of the $\lambda_l(t_j)$ over subjects that indexed by l in the risk set at time t_j, $R(t_j)$. Therefore, the conditional probability above can be rewritten as

$$\frac{\lambda_j(t_j)}{\sum_{l \in R(t_j)} \lambda_l(t_j)} = \frac{\exp(\boldsymbol{\beta}' \boldsymbol{x_j})}{\sum_{l \in R(t_j)} \exp(\boldsymbol{\beta}' \boldsymbol{x_l})} \qquad (1.8)$$

where $\lambda_0(t_j)$ in the numerator and denominator cancels out.

Taking the product of (1.8) over the r event times, we get

$$L^{PL}(\beta) = \prod_{j=1}^{r} \frac{\exp(\boldsymbol{\beta}' \boldsymbol{x_j})}{\sum_{l \in R(t_j)} \exp(\boldsymbol{\beta}' \boldsymbol{x_l})}$$

When censoring is present in the data, the partial likelihood can then be expressed in the form

$$L^{PL}(\beta) = \prod_{j=1}^{n} \left[\frac{\exp(\boldsymbol{\beta}' \boldsymbol{x_j})}{\sum_{l \in R(t_j)} \exp(\boldsymbol{\beta}' \boldsymbol{x_l})} \right]^{\delta_j} \tag{1.9}$$

Another derivation for the partial likelihood function is as follows. Assuming independent event and censoring times, the full likelihood function for censored data is given by

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{n} [\lambda_j(y_j)]^{\delta_j} S_j(y_j)$$

$$= \prod_{j=1}^{n} [\lambda_0(y_j) \exp(\boldsymbol{\beta}' \boldsymbol{x_j})]^{\delta_j} \exp\left[- \int_0^{y_j} \lambda_0(s) \exp(\boldsymbol{\beta}' \boldsymbol{x_j}) ds \right] \tag{1.10}$$

By multiplying and dividing (1.10) by the term $\left[\sum_{l \in R(t_j)} \lambda_0(y_j) \exp(\boldsymbol{\beta}' \boldsymbol{x_l})\right]^{\delta_j}$, and $l \in R(t_j)$ means $t_l \geq t_j$, we get

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{n} \left[\frac{\exp(\boldsymbol{\beta}' \boldsymbol{x_j})}{\sum_{l \in R(t_j)} \exp(\boldsymbol{\beta}' \boldsymbol{x_l})} \right]^{\delta_j}$$

$$\times \left[\sum_{l \in R(t_j)} \lambda_0(y_j) \exp(\boldsymbol{\beta}' \boldsymbol{x_l}) \right]^{\delta_j} \exp\left[- \int_0^{y_j} \lambda_0(s) \exp(\boldsymbol{\beta}' \boldsymbol{x_j}) ds \right] \tag{1.11}$$

Cox (1972) argued that the first term in (1.11) takes into consideration the ordering of the events and does not make use of the actual event times. So that last bit of information is what is dropped. Thus, the estimates of $\boldsymbol{\beta}$ can be obtained from the first part of (1.11) which is the partial likelihood function defined in (1.9).

This partial likelihood can be maximized using the Newton-Raphson procedure. In the case of ties, approximations to the partial likelihood such as those proposed by Breslow (1974) and Efron (1977) are needed.

The key difference between the parametric and semi-parametric PH models is that instead of leaving the baseline hazard completely arbitrary in the semi-parametric PH approach, the baseline hazard is assumed to follow a specific distribution in the parametric PH approach. Further, the estimation in the parametric PH model is based on the full maximum likelihood instead of the partial likelihood used in the Cox PH model.

1.4.2. Accelerated failure time (AFT) model

The accelerated failure time model is an alternative to the PH model for the analysis of survival data. Under AFT models, the effect of the explanatory variables is measured directly on the survival time instead of on the hazard function as in PH models. This allows for easier interpretation of the results because the predictors affect the mean survival time through the regression parameters. The AFT model takes the form

$$\lambda(t) = \lambda_0\big(t \exp(\boldsymbol{\beta}'\boldsymbol{x_j})\big) \exp(\boldsymbol{\beta}'\boldsymbol{x_j}) \qquad (1.12)$$

where the change from t to $t\exp(\boldsymbol{\beta}'\boldsymbol{x_j})$ represents acceleration or deceleration depending whether $\exp(\boldsymbol{\beta}'\boldsymbol{x_j})$ is greater or smaller than 1. Only PH models with an exponential or a Weibull baseline survival time distribution belong to both the PH and AFT families.

The log-linear formulation corresponding to the AFT model with respect to time is given by

$$\ln T_j = \mu + \boldsymbol{\beta}'\boldsymbol{x_j} + \sigma\varepsilon_j \tag{1.13}$$

where μ is an intercept, σ is a scale parameter, and ε_j is the random error term for the j^{th} subject which is assumed to follow a certain distribution. This log-linear form representation is adopted by most of the software packages.

1.4.3. Proportional odds model

The proportional odds model is structurally similar to the PH model, and may be used in similar situations (Bennett, 1983). In the situations where the predictor effect vanishes with time, the proportional odds model may be more appropriate than the PH model. Similar to the PH model, the odds function is assumed to be a product of baseline odds and an exponential function of the predictors. The proportional odds model is given by

$$\mu(t) = \mu_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{x_j}) \tag{1.14}$$

where $\mu(t) = F(t)[1 - F(t)]^{-1}$ is the odds function at time t, $\mu_0(t) = F_0(t)[1 - F_0(t)]^{-1}$ is the baseline odds function at time t of a subject with $\boldsymbol{x_j} = \mathbf{0}$. Likelihood-based estimation can be used for inference on regression parameters $\boldsymbol{\beta}$.

1.4.4. Additive risks model

Another alternative model to the PH model is the additive risks model (Aalen, 1980; 1989). Unlike the PH model, the additive risk model assumes that the hazard function for the j^{th} subject associated with a set of predictors $\boldsymbol{x_j}$ is the sum of a baseline hazard function $\lambda_0(t)$ and a regression function of explanatory variables $\boldsymbol{x_j}$, that is

$$\lambda(t) = \lambda_0(t) + \boldsymbol{\beta}'\boldsymbol{x_j} \tag{1.15}$$

where $\lambda_0(t)$ is the baseline hazard. The regression parameters $\boldsymbol{\beta}$ are allowed to be functions of time, so that the effect of predictors may vary over time. Model parameters can be estimated using a least-squares technique (Huffler and McKeague, 1991).

1.5. Time-dependent predictors and coefficients

In many studies with time-to-event outcome, individuals are monitored for the duration of the study. Within this period, the values of certain explanatory variables may be recorded at selected periodic time points (i.e. $\boldsymbol{x_j}$ is a function of t). For example, predictors related to management practices may be recorded at regular intervals. These variables are known as time-dependent or time-varying predictors. Another type of time dependence is when the effect of certain predictors changes over time (i.e. $\boldsymbol{\beta}$ is a function of t), this is referred to as time-dependent coefficients or time-varying effects. To deal with time-dependent variables and coefficients in PH models the dataset has to be set up in a counting-process format (Grambsch and Therneau, 1994). In the counting-process setup, data for each subject are identified by the triple: $N(t)$, $\delta(t)$ and $\boldsymbol{x}(t)$ where $N(t)$ is the number of events that occurred in the time interval $(0, t]$ for a subject j; $\delta(t)$ indicates the event status; and $\boldsymbol{x}(t)$ is a vector of predictors for the subject at time t. The

path of $N(.)$ is a step function with jumps at the event times and $N(0) = 0$. The data for each subject are then represented by multiple records, each identifying a time interval $(t_k, t_{k+1}]$, the predictor values on that interval, and the event status. Finally, the time dependence of predictors and effects can be accounted for by fitting, for example, a PH model to the counting-process formatted data (Therneau and Crowson, 2013).

1.6. Hierarchical data structure

Clustering in the data is natural in most observational and experimental studies in veterinary epidemiology. For example, in the lameness dataset described in Section 1.2.1 we have animals clustered by herds. Animals within the same herd are more alike than animals from different herds, in the sense that animals within the same herd share the experience of being in the same environment and similar genetics (e.g. food, facilities, and management). Another type of clustering in survival data is where we have repeated events in the same subject, for example repeated CM episodes in dairy cows (Schukken et al., 2010). Such similarity may lead to within cow homogeneity over time (Dohoo et al., 2009; chapter 21). We talk about a hierarchical data structure when each unit at the lower level is nested in a single unit of the higher levels. In addition, any two observations in the same unit must remain together (i.e. in the same unit) at all higher levels. Two-level hierarchical survival data, e.g. animals within herds, are common in human and veterinary epidemiologic studies (e.g. Stryhn and Christensen, 2013; Hanagal and Dabade, 2013). Analysis of survival data with more than two levels of hierarchy have been reported in the human epidemiology literature (e.g. Sastry, 1997; Shin and Lu, 2007); but are rarely reported in veterinary epidemiological research even though such structures are commonly encountered in the field. For example, the CBMRN dataset

described in Section 1.2.2 includes multiple lactations nested in cows and cows located in different herds.

In some instances, the data structure is not perfectly hierarchical. For instance, the structure of the calf mortality dataset is more complex than what was described in Section 1.2.3. We have calves from different herds, and these herds located in different ecologic regions and serviced by multiple veterinary clinics. Therefore, in addition to calves being hierarchically nested within herds, herds are cross-classified by ecologic regions and veterinary clinics. This special structure corresponds to a 3-level cross-classified and multiple membership data structure (Browne et al., 2001).

It is important to understand the implications of ignoring clustering or otherwise inadequately accounting for a hierarchical data structure in the statistical analysis, beyond the obvious fact that the independence assumption is being violated. Biases may occur in both regression coefficients (in particular if groups have confounding effects) and in their standard errors. It is well-known that standard errors for group-level predictors will be underestimated if clustering exists and is ignored. Thus, the researcher may be more likely to conclude that an effect (e.g. a hazard ratio or a difference between group means) is statistically significant regardless of whether an effect is actually present in the population.

Different approaches have been applied to account for hierarchical survival data structure. One of these approaches is to combine multiple hierarchical levels of analysis in a single comprehensive model by including the information from each level of the hierarchy in the data. This allows researchers to specify predictors at different levels and

apply other features such as random slopes and contextual effects. In the following section, a brief review of different approaches for accounting for data structure in survival data is given.

1.7. Approaches for modeling two-level hierarchical survival data

In this section, the existing approaches for modeling hierarchical data structure for time-to-event outcomes are discussed. Based on a Cox PH model, three extensions can be considered for modeling a two-level data structure: a model including the top level as fixed effects, a model stratified by the top level, and a model incorporating the top level as random effects. Other approaches, such as copula models, have been also suggested for correlated event times.

In the following subsections, hierarchical survival data clustered by G different groups are considered. The subject j ($j = 1, \dots, n_i$) in group i ($i = 1, \dots, G$) is either observed from time zero to an event time T_{ij} or to a right censoring time C_{ij} independent of T_{ij}. As described in Section 1.2.2, let $Y_{ij} = \min(T_{ij}, C_{ij})$ be the observed time and δ_{ij} be the event indicator. For each subject, we also observe the vector of explanatory variables $\boldsymbol{x_{ij}}$.

1.7.1. Fixed effect model approach

One simple approach to dealing with hierarchical survival data is to include the group as a categorical fixed effect in the model. By arbitrarily setting one group as reference, a model with $G - 1$ dummy variables representing the group variable can be estimated using the Cox PH model methodology. This approach can be worthwhile for a small number of groups and no group-level predictors. However, in the case of a large number

of groups and a small group size, the parameter estimation might become unstable and the efficiency of estimates may be affected (Dohoo et al., 2009). For instance, the log-likelihood sometimes converges, even when the estimate of a specific group diverges, this phenomenon occurs when all event times in a specific group are smaller or larger than event times in the other groups (Legrand, 2010; p. 21). A monotone likelihood has been proposed to solve this problem (Heinze and Schemper, 2001). In addition to the aforementioned issues, there are other drawbacks to the fixed effect model approach. First, one cannot estimate any between-group fixed effects as they will be absorbed into the group effects. Second, any inferences based on this approach are specific to the actual groups, not to the general population of groups, one would often want conclusions to refer to.

1.7.2. Stratified effect model approach

Another approach to deal with clustering in survival data is to fit a Cox PH model stratified by groups. This procedure allows a specific unspecified baseline hazard for each group. The stratified Cox PH model can be written as,

$$\lambda_{ij}(t) = \lambda_{0i}(t)\exp(\boldsymbol{\beta}'\boldsymbol{x_{ij}}) \qquad (1.16)$$

where $\lambda_{0i}(t)$ is the baseline hazard function for group i at time t. The proportional hazards assumption of this model is not assumed across groups, but only within each group. Model (1.16) can be estimated by combining the partial likelihood for each group as follows,

$$L_{str}(\boldsymbol{\beta}) = \prod_{i=1}^{G} \prod_{j=1}^{n_i} \left[\frac{\exp(\boldsymbol{\beta}' \boldsymbol{x_{ij}})}{\sum_{l \in R(t_{ij})} \exp(\boldsymbol{\beta}' \boldsymbol{x_{ij}})} \right]^{\delta_{ij}} \quad (1.17)$$

The asymptotic theory for the stratified Cox model is valid for frameworks where n_i remains bounded as G increases, where the group size increases with G, and where G is fixed and n_i increases (Glidden and Vittinghoff, 2004). However, applying this approach leads to discarding an abundant amount of information. Specifically, no between-group comparisons can be made, and all information related to the predictor effect is based on within-group comparisons. For a fixed sample size, the loss of information increases with G (Glidden and Vittinghoff, 2004). The groups that contain no events and/or only one-predictor level do not contribute to the model estimates leading to inefficient estimates. Also, when the aim is to quantify the variation in the (baseline) hazard between groups, the stratified Cox model is no longer of interest, as it does not provide such information.

1.7.3. Shared frailty (random effects) model approach

Similar to the fixed effect model, the random effects or frailty model assumes that group effects act proportionally on the baseline hazard. However, the random effects or frailty model treats the group effects as random effects, i.e. as a sample from a certain probability distribution. The model is given by

$$\lambda_{ij}(t|u_i) = \lambda_0(t)\, u_i \exp\left(\boldsymbol{\beta}' \boldsymbol{x_{ij}}\right) \quad (1.18)$$

where the frailty term u_i acts multiplicatively on the baseline hazard, and represents the effect of unobserved factors. The frailties u_i are assumed independently and identically distributed with unity mean (in order to make the average hazard identifiable) and

unknown variance θ. Large values of θ indicate a closer relationship between the subjects of the same group and greater heterogeneity among the groups. The model defined in (1.18) is known as a shared frailty model (Therneau and Grambsch, 2000; Duchateau and Janssen, 2008; Wienke, 2010), in the sense that unmeasured characteristics, such as genetic information or common environmental exposures are correlated by the subjects within the same group or cluster. In this way, these unobserved factors could influence time to the event of interest. An alternative formulation for a shared frailty model is

$$\lambda_{ij}(t|b_i) = \lambda_0(t)\exp\left(b_i + \boldsymbol{\beta}'\boldsymbol{x_{ij}}\right) \tag{1.19}$$

with $b_i = \ln u_i$ is the random effect.

To distinguish between the frailties and the random effects, we denote throughout the thesis by u_i the frailty and by b_i the random effect, and by θ and σ^2 the variance of frailty and random effects, respectively.

1.7.4. Copula model approach

Another approach for modeling hierarchical survival data is through a copula model. This approach was originally introduced to be used for datasets with small and equal group sizes and it has been recently developed to allow for varying group sizes. In the veterinary field, Goethals et al. (2008) studied a copula model for bivariate survival data on diagnosis of fracture healing in dogs and compared it with a shared frailty model. Massonnet et al. (2009) applied a copula model to the infection times in the four udder

quarters of dairy cows. The survival copula is a function that links marginal survival functions to generate a joint survival function, i.e.,

$$S_i(t_1, t_2) = C_\varphi\{S_{i1}(t_1), S_{i2}(t_2)\} \quad (1.20)$$

for a bivariate distribution copula function C_φ defined as $C_\varphi: [0,1]^2 \rightarrow [0,1]: (\nu_1, \nu_2) \rightarrow C_\varphi(\nu_1, \nu_2)$ parameterized by φ. Parametric copula models can be estimated using a one-stage estimation approach, while a two-stage approach is needed for semiparametric copula models. In semiparametric copula models, the marginal survival functions are estimated first, and then the copula parameter φ is estimated by maximizing a log-likelihood function after replacing the marginal survival functions by their estimates obtained in the first stage. The class of Archimedean copulas is the most considered class of survival copulas. Detailed discussion on copulas can be found in Li (2000) and Nelsen (2006).

1.8. Frailty modeling approaches for more complex hierarchical data structures

1.8.1. Random slope model

In the shared frailty model, the unobserved heterogeneity is not captured by the predictors and is assumed to be the same within each group. Further, the shared frailty model can only induce positive relationships within the group, although event times may be negatively associated in some situations (Xue and Brookmeyer, 1996). A model with random group and random predictor effects has been suggested to overcome the limitations of the shared frailty model (Ripatti and Palmgren, 2000). The two-level hierarchical Cox model with random baseline hazard and random slope is given by:

$$\lambda_{ij}(t|b_{i0}, b_{i1}) = \lambda_0(t)\exp\left(b_{i0} + (b_{i1} + \beta)x_{ij}\right) \quad (1.21)$$

where b_{i0} and b_{i1} are jointly distributed and represent, respectively, the random group and the random slope, and x_{ij} is the observed predictor for subject j in group i. The term random slope is most meaningful for a quantitative predictor x where beta corresponds to a slope. If x is dichotomous (say treatment), the random "slope" is a random treatment effect, and could also be understood as a random interaction between treatment and groups. For simplicity, the term random slope will be used throughout the thesis, even in cases where x is not quantitative.

1.8.2. Nested frailty model

When event times are clustered at several hierarchical levels such as herds and geographic regions, nested frailty models can be used to account for the hierarchical data structure by including nested frailty terms, where by each frailty term represents a level of clustering and acts multiplicatively on the baseline hazard (Sastry, 1997). For observations clustered by SG subgroups nested in G groups, the nested frailty model can be written as,

$$\lambda_{ijk}(t) = \lambda_0(t)\, u_i u_{ij} \exp\left(\boldsymbol{\beta}' \boldsymbol{x_{ijk}}\right) \quad (1.22)$$

with u_i and u_{ij} are the nested frailties that correspond to the group and subgroup levels, respectively. Model (1.22) can be rewritten as

$$\lambda_{ijk}(t) = \lambda_0(t)\exp\left(b_i + b_{ij} + \boldsymbol{\beta}' \boldsymbol{x_{ijk}}\right) \quad (1.23)$$

where the nested random effects b_i and b_{ij} represent, respectively, the deviation of the i^{th} group, and the j^{th} subgroup of the i^{th} group from the overall log baseline hazard.

1.8.3. Multiple membership model

As in other types of datasets, multiple membership structure may arise in hierarchical survival data where the units of a lower hierarchical level are members of multiple higher level units simultaneously (Browne et al., 2001). Multiple membership models have been proposed to account for such data structure (Browne et al., 2001; Fielding and Goldstein, 2006) by using weights for the units that occur in a multiple membership relation. For example, a hospital patient may be treated by several nurses and each nurse will then have an effect on the patient's progress (Browne et al., 2001). In this example, we have patients, each cared for by single or multiple nurses resulting in a 2-level data structure, where some of the lower-level units are in a multiple membership with the top level units, the multiple membership model can be written in a Cox model formulation as,

$$\lambda_j\left(t|(b_l^{(2)})_{l\in group(j)}\right) = \lambda_0(t)\exp\left(\sum_{l\in group(j)} w_{jl}^{(2)} b_l^{(2)} + \boldsymbol{\beta}'\boldsymbol{x_j}\right) \qquad (1.24)$$

where the term $\sum_{l\in group(j)} w_{jl}^{(2)} b_l^{(2)}$ involves a set of random effects $b_l^{(2)}$ at the second level and weights $w_{jl}^{(2)}$ assigned to each second-level unit for their group membership with $\sum_{l\in group(j)} w_{jl} = 1$. To our knowledge, this special type of models has only been applied in the context of linear and generalized linear models, but it has not been used with survival data.

1.9. Frailty and random effects distributions

Various frailty and random effects distributions have been used for modeling hierarchical survival data. Due to software availability, some of these distributions are more commonly applied in the area than others. In practice, the gamma and log-normal distributions with unity-mean are the most applied frailty distributions. Other choices for frailty distributions, such as inverse Gaussian, positive stable (Hougaard, 1995), power variance function (Aalen, 1988), and compound Poisson (Aalen, 1992) have also been used in the literature. See Hougaard (2000) and Duchateau and Janssen (2008) for in-depth discussions.

1.9.1. Gamma frailty and log-gamma random effects distributions

When the frailties $u_1, \dots, u_G$ follow a gamma distribution with the same shape and inverse scale parameters of $1/\theta$, their density function is given by

$$f_U(u) = \frac{\theta^{-\theta^{-1}}}{\Gamma(\theta^{-1})} u^{\theta^{-1}-1} \exp(-\theta^{-1}u), \quad u \geq 0 \qquad (1.25)$$

Therefore, the distribution of U has a mean of 1 and variance of θ, and the random effects $b_1 = \ln u_1, \dots, b_G = \ln u_G$ have a log-gamma distribution with density function given by

$$f_B(b) = \frac{\theta^{-\theta^{-1}}}{\Gamma(\theta^{-1})} \exp(\theta^{-1}b - \theta^{-1}\exp(b)), \quad -\infty < b < \infty \qquad (1.26)$$

The variable B has a mean of $\Psi(1/\theta) + \log(\theta)$ and a variance of $\sigma^2 = \Psi'(1/\theta)$, where $\Psi(.)$ and $\Psi'(.)$ are the digamma and trigamma functions, respectively. The log-gamma distribution with mean zero and variance 0.5 for random effects is presented in the left of

Figure 1.1 and the one-parameter gamma distribution with variance 0.5 for frailty is depicted in the right of Figure 1.1.

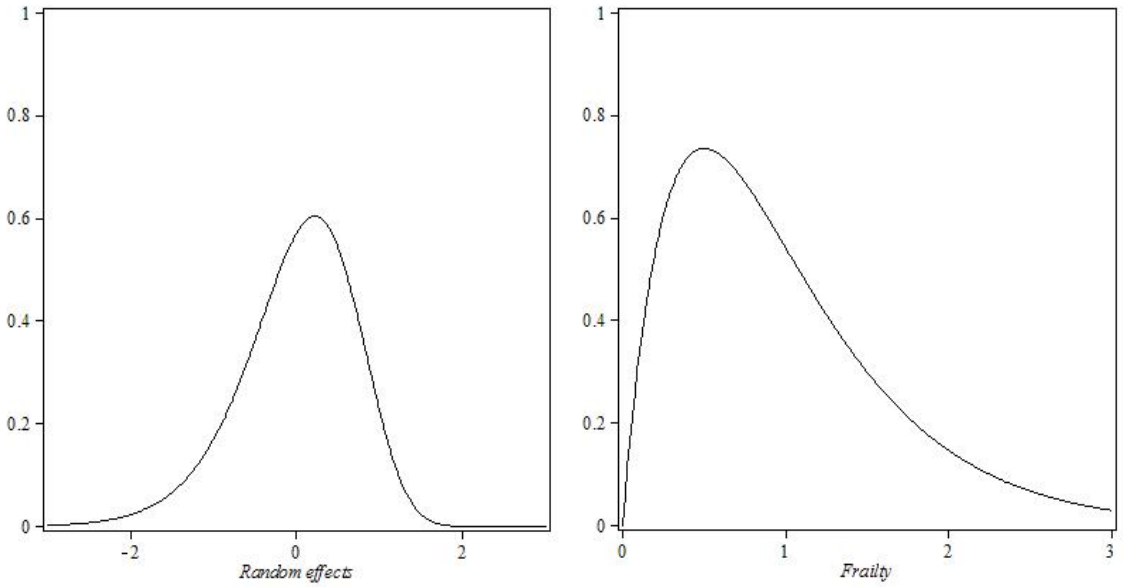

Figure 1.1. The log-gamma distribution with mean zero and variance 0.5 for random effects (left) and the gamma distribution with mean 1 and variance 0.5 for frailty (right). Note that the two density curves are not representations of the same distribution on logarithmic and exponential scales, respectively, because the mean value restrictions are incompatible (e.g., a mean of 1 on frailty scale does not lead to a mean of 0 after log-transformation).

1.9.2. Normal random effects and log-normal frailty distributions

For the random effects $b_1, \dots, b_G$ that follow a normal distribution with mean zero and variance σ^2, the density function is given by

$$f_B(b) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-b^2/2\sigma^2), \quad -\infty < b < \infty \qquad (1.27)$$

Thus, the frailties $u_1 = \exp(b_1), \dots, u_G = \exp(b_G)$ follow a log-normal distribution with a density function,

$$f_U(u) = \frac{1}{u\sigma\sqrt{2\pi}}\exp(-(\ln u)^2/2\sigma^2), \quad u \geq 0 \tag{1.28}$$

The mean and the variance of U are then $\exp(\sigma^2/2)$ and $\exp(\sigma^2)(\exp(\sigma^2) - 1)$, respectively. The zero-mean normal distribution for random effects and unity-mean log-normal distribution for frailty, both with variance of 0.5 are presented in the left and the right sides of Figure 1.2, respectively.

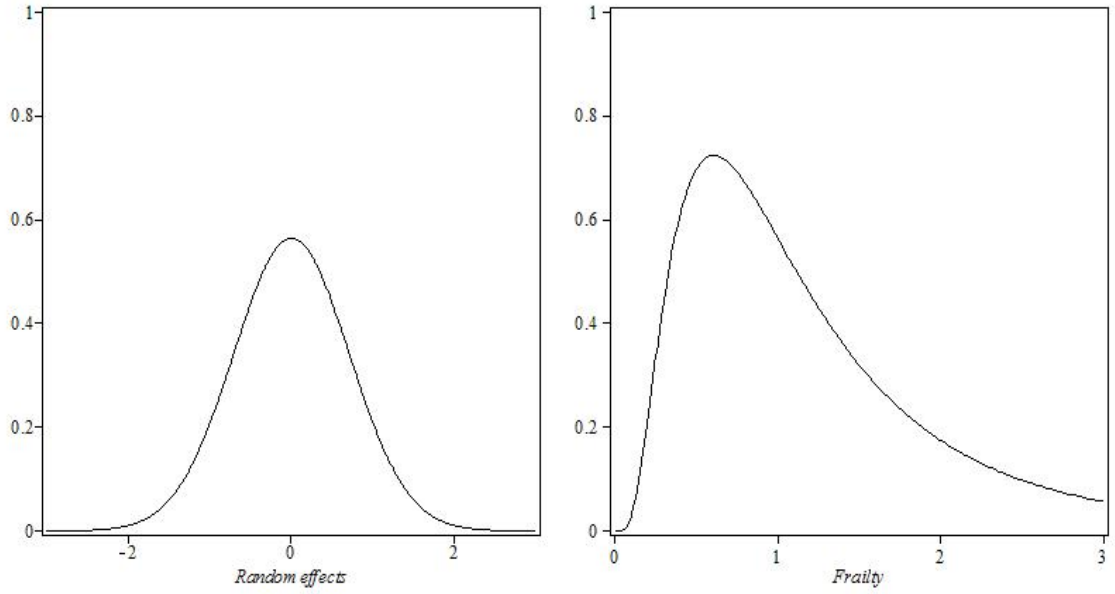

Figure 1.2. The normal distribution with mean zero and variance 0.5 for random effects (left) and the log-normal distribution with mean 1 and variance 0.5 for frailty (right). As noted above, the two distributions are not transformations of each other.

1.9.3. Laplace random effects and log-Laplace frailty distributions

If the random effects $b_1, \ldots, b_G$ have a Laplace distribution with a location parameter of 0 and scale parameter τ, then the density function is given by

$$f_B(b) = \frac{1}{2\tau}\exp\left(\frac{|b|}{\tau}\right), \quad -\infty < b < \infty \tag{1.29}$$

The variance of the distribution is $2\tau^2$. Figure 1.3 shows the probability density plot of Laplace distribution with mean zero and variance 0.5.

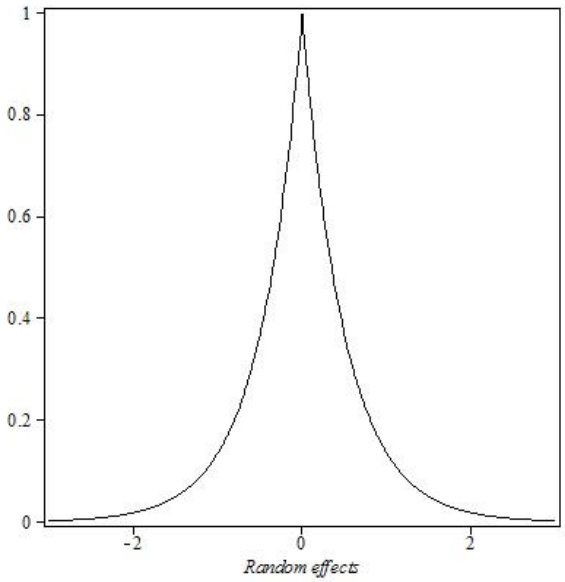

Figure 1.3. The Laplace distribution with mean zero and variance 0.5 for random effects.

The frailties $u_1 = \exp(b_1), \ldots, u_G = \exp(b_G)$ have a log-Laplace density function,

$$f_U(u) = \frac{1}{2\tau u}\exp\left(\frac{|\ln u|}{\tau}\right), \quad u \geq 0 \tag{1.30}$$

The random effects densities of normal, log-gamma, and Laplace distributions with mean zero and variance 0.5 are presented in Figure 1.4.

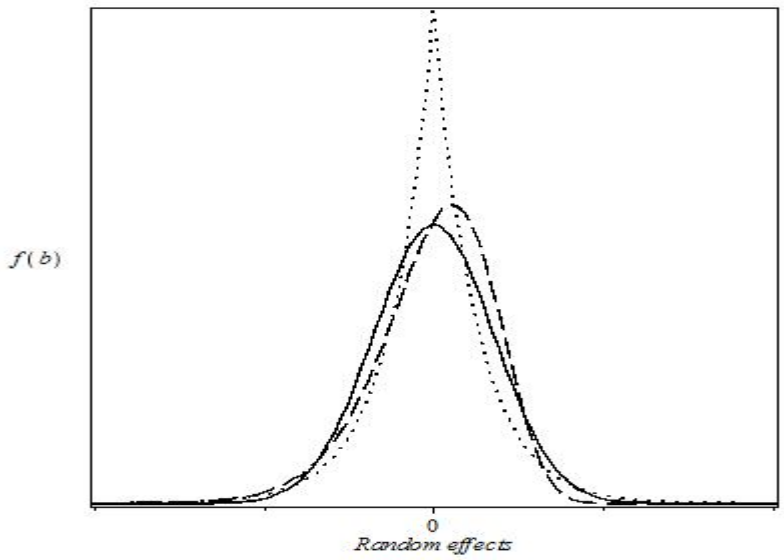

Figure 1.4: The normal (solid), log-gamma (dash), and Laplace (dot) probability densities with mean zero and variance 0.5 for random effects.

1.10. Estimation methods for semiparametric hierarchical survival models

The focus here and throughout this thesis is on semiparametric estimation approaches where the form of the baseline hazard is left completely arbitrary. Parameter estimation in hierarchical survival (frailty) models has been a topic for intensive research over the past few years, and numerous estimation approaches have been proposed. Some of these estimation approaches have been reviewed and discussed by other authors, for example, Clayton (1988), Duchateau et al. (2002), Cortiñas et al. (2007), Duchateau and Janssen (2008), Hanagal (2011), and Hirsch and Wienke (2012). A summary of some approaches are presented in Table 1.1. Further, current implementations of these models are shown in Table 1.2.

Table 1.1. Summary of estimation approaches for semiparametric hierarchical proportional hazards models and their applicability to different designs, each identified by the key/first reference describing the method.

Estimation approach	Model used		
	Shared frailty	**Random slope**	**Nested frailty**
EM algorithm	Klein (1992) for gamma frailty.	Cortiñas and Burzykowski (2005) for normal random effects	
Penalized partial likelihood	McGilchrist (1993) for normal random effects. Therneau and Grambsch (2000) for gamma frailty.	Yamaguchi and Ohashi (1999) for normal random effects.	Yau (2001)
Penalized partial likelihood with Laplace approximation	Ripatti and Palmgren (2000) for normal random effects.	Ripatti and Palmgren (2000) for normal random effects.	Ripatti and Palmgren (2000) for normal random effects.
Penalized full likelihood	Rondeau et al. (2003) for gamma and log-normal frailty.	Rondeau et al. (2008) for normal random effects.	Rondeau et al. (2006) for gamma frailty.
Hierarchical likelihood	Ha et al. (2001) for gamma frailty and normal random effects.		Ha et al. (2007) for normal random effects.
Poisson maximum likelihood	Ma et al. (2003) for gamma frailty and normal random effects.	Ma et al. (2003) for normal random effects.	Ma et al. (2003) for gamma frailty and normal random effects.
Bayesian	Clayton (1991) for normal random effects.		
Monte Carlo EM	Vaida and Xu (2000) for normal random effects.	Vida and Xu (2000) for normal random effects.	Gamst et al. (2009) for normal random effects.
Bayesian with Laplace approximation	Ducrocq and Casella (1996) for normal random effects.	Legrand et al. (2005) for normal random effects.	

In addition to the data structure, the choice of the estimation approach may depend on the desired frailty or random effects distribution, but in practice the choice is mainly determined by the availability of appropriate software. Mathematically, closed forms for the survival and hazard functions can be obtained under a gamma frailty distribution, but when a normal distribution is assumed for random effects, explicit expressions for these functions do not exist. In this case, approximations are needed to overcome the problem.

In the following sections, a brief overview is given of some of the semiparametric estimation approaches for different hierarchical PH models. The review is limited to the penalized partial likelihood approach with both gamma and log-normal frailty distributions and to the estimation approaches used in this thesis, namely, the penalized full likelihood approach (Rondeau et al., 2003), the Poisson maximum likelihood approach (Ma et al., 2003; Feng et al., 2005), the Bayesian approach (Clayton, 1991; 1994), and the penalized partial likelihood with Laplace approximation (Ripatti and Palmgren, 2000). The discussion is based on shared frailty model estimation unless otherwise stated.

Table 1.2. Summary of existing software for semiparametric models of hierarchical survival data.

Software implementation	Model design	Estimation method	Description
coxph	Gamma, lognormal, and log-t shared frailty.	Penalized partial likelihood.	R functions in survival package.
coxme	Shared, random slope, and nested random effects with a normal distribution.	Penalized partial likelihood with Laplace approximation.	R functions in coxme package.
frailtyPenal	Gamma shared and nested frailty.	Penalized full likelihood with splines.	R functions in frailtypack package.
additivePenal	Gaussian random slope.	Penalized full likelihood with splines.	R functions in frailtypack package.
phmm	Shared and random slope with a normal distribution.	Monte Carlo EM	R functions in phmm package.
frailtyHL	Gamma and lognormal shared and nested frailty.	Hierarchical likelihood	R functions in frailtyHL package.
stcox (with option shared)	Shared gamma frailty	Penalized partial likelihood.	Stata command
SPGAM	Shared gamma frailty	EM algorithm	SAS macro
SPLN3	Shared lognormal frailty.	EM algorithm	SAS macro
Gamfrail	Shared gamma frailty	EM algorithm	SAS macro
The survival kit	Shared and random slope with a normal distribution.	Bayesian with Laplace approximation.	Package of Fortran programs.
Poisson GLMM	Shared, random slope, and nested random effects with either normal or gamma distributions.	Poisson maximum likelihood.	Any GLMM software.
Bayesian	Shared, random slope, and nested random effects.	MCMC techniques.	Any Bayesian software for example WinBUGS.

1.10.1. The penalized partial likelihood approach

This approach is an extension of the best linear unbiased predictor (BLUP) to be used for multivariate survival data as described in McGilchrist and Aisbett (1991) and McGilchrist (1993). The approach is implemented in coxph function of R software and stcox command of Stata and became widely used in the area of multilevel survival

analysis. As described in Duchateau and Janssen (2008), the penalized partial log-likelihood for model (1.19) is given by

$$l^{PPL}(\beta, \sigma^2, b) = l^{PL}(\beta, b) - l^{pen}(\sigma^2, b) \tag{1.31}$$

where $l^{PL}(\beta, b)$ is the logarithm of (1.9) with b as another set of parameters and defined as

$$l^{PL}(\beta, b) = \sum_{i=1}^{G} \sum_{j=1}^{n_i} \delta_{ij} \left[\left(b_i + \beta' x_{ij} \right) - \ln \sum_{(p,q) \in R(t_{ij})} \exp(b_p + \beta' x_{pq}) \right]$$

and $l^{pen}(\sigma^2, b) = -\sum_{i=1}^{G} \ln f_b(b_i)$ is the penalty function. This penalty function reduces the penalized partial likelihood by shrinking the random effects towards the mean value zero.

For random effects $b_i, i = 1, \dots, G$, with a zero-mean normal distribution, the penalty term takes the form,

$$l^{pen}(\sigma^2, b) = \frac{1}{2} \sum_{i=1}^{G} \left[\frac{b_i^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]$$

The maximization process of the penalized partial log-likelihood consists of an inner and an outer loop. For provisional value of σ^2, $l^{PPL}(\beta, \sigma^2, b)$ is maximized for β and b in the inner loop using the Newton-Raphson procedure. In the outer loop, the residual maximum likelihood (REML) estimator for σ^2 is obtained using the best linear unbiased predictors (BLUPs) for b. This process is iterated until achieving convergence.

For frailties $e^{b_i}, i = 1, \dots, G$, following a gamma distribution $\Gamma(\theta^{-1}, \theta^{-1})$, the penalty function is

$$l^{pen}(\theta, b) = -\sum_{i=1}^{G}\left(\frac{b_i - e^{b_i}}{\theta}\right) - G\left(\frac{\ln\theta}{\theta} - \ln\Gamma(\theta^{-1})\right)$$

As in the maximization process of the random effects with a normal density, inner and outer loops are used to maximize the penalized partial likelihood. The inner loop is the same as the one described in the normal density case with exception of the penalty term that is determined based on the gamma density. For the outer loop, a REML estimator for θ cannot be obtained as in the case of a normal random effect distribution; therefore, a profiled version of the following Klein's marginal log-likelihood (Klein, 1992) is maximized for θ:

$$\begin{aligned} l(\lambda_0(.), \beta, \theta) = &\sum_{i=1}^{G}(D_i \ln\theta - \ln\Gamma(\theta^{-1}) + \ln\Gamma(\theta^{-1} + D_i)) \\ &- \sum_{i=1}^{G}(\theta^{-1} + D_i)\ln\left[1 + \theta\sum_{j=1}^{n_i}\Lambda_0(y_{ij})\exp(\beta' x_{ij})\right] \\ &- \sum_{i=1}^{G}\sum_{j=1}^{n_i}\delta_{ij}[\ln(\lambda_0(y_{ij}) + \beta' x_{ij}] \end{aligned} \quad (1.32)$$

where $D_i = \sum_{j=1}^{n_i}\delta_{ij}$ are the observed events in the i^{th} group.

The algorithm is iterated until convergence. Detailed explanations and an excellent review are provided by Duchateau and Janssen (2008).

Yamaguchi and Ohashi (1999) extended the penalized partial likelihood approach to estimate treatment-by-centre (random slope) in a multicentre trial. Also, the approach was extended by Yau (2001) to estimate a three-level hierarchical survival model.

1.10.2. The penalized full likelihood approach

The approach was proposed by Rondeau and Gonzalez (2003) to fit a shared gamma frailty model by using splines to model the baseline hazard function. Because of the splines, the procedure is similar to the parametric approach using a piecewise constant baseline hazard. However, this approach is much more flexible than the classical parametric approach. When the number of pieces becomes large, the approach is considered semiparametric because the model shows similar flexibility as the semiparametric model. Rondeau and Gonzalez (2003) used the penalized log-likelihood,

$$l_{PFL}(\lambda_0(.),\beta,\theta) = l(\lambda_0(.),\beta,\theta) - v\int_0^{\infty}[\lambda_0''(t)]^2 dt \qquad (1.33)$$

where $l(\lambda_0(.),\beta,\theta)$ is the marginal log-likelihood defined in (1.32), $\lambda_0''(t)$ is the second derivative of the baseline hazard, and v is a positive smoothing parameter that controls the trade-off between the data fit and the smoothness of the function $\lambda_0(.)$.

Rondeau and Gonzalez (2003) used cubic M-splines (Ramsay, 1988) that are easy to integrate and differentiate, the second derivative of $\lambda_0(.)$ being approximated by a linear combination of polynomial terms. Such an approximation reduces the number of parameters but still allows for flexible shapes of hazard functions. The approximation error can be made as small as desired by increasing the number of knots. The smoothing parameter can be fixed by the user or automatically estimated by maximizing a likelihood

cross-validation criterion for the Cox model (Joly et al., 1998; Rondeau and Gonzalez, 2003). The log-likelihood in (1.33) is maximized by the robust Marquardt algorithm (Marquardt, 1963), which is a combination between a Newton-Raphson and a steepest descent algorithm. The approach is implemented in the frailtypack package for R software (Rondeau et al., 2012).

Rondeau et al. (2006) extended the approach to allow for estimating models with two nested frailties assuming a gamma distribution at the lower level and a normal distribution on the log scale at the higher level, and a further extension was added by Rondeau et al. (2008) for estimating two-level hierarchical models with a random slope.

1.10.3. The Poisson maximum likelihood

The similarity of a Cox PH model with a Poisson regression model has been known since 1980 (Whitehead, 1980). This similarity can be carried over to the Cox PH model with random effects (Ma et al., 2003; Feng et al., 2005; Rabe-Hesketh and Skrondal, 2012). Using available software for generalized linear mixed models (GLMMs), model (1.19) can be estimated through a Poisson GLMM framework after expanding the data into the counting-process format as follows:

Ignoring random effects and using the notation of Section 1.3, the contribution of subject j from group i to the likelihood is

$$L_{ij}(\beta) = \left[\lambda_{ij}(t)\right]^{\delta_{ij}} \exp\left(-\int_0^t \lambda_{ij}(s)\, ds\right) \qquad (1.34)$$

In the piecewise exponential model, the baseline hazard function is assumed to be piecewise constant, with $\lambda_0(s) = \lambda_{0q},\ t_{q-1} \leq s < t_q,\ q = 1, \dots, Q$ and $l_q = t_q - t_{q-1}$ is the interval length. Proceeding this way, the baseline hazard function can be estimated nonparametrically by letting the t_q $(q = 1, \dots, Q)$ correspond to the observed failure times and Q be the total number of distinct observed failure times in the study (previously denoted by r). As described in Clayton (1988), the contribution of the subject j that experienced the event or was censored in the interval k_j to the likelihood is

$$L_j(\beta) = \left[\lambda_{0q}\exp\left(\boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]^{\delta_{ij}} \exp\left[-\sum_{q=1}^{k_j} \lambda_{0q} l_q \exp\left(\boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right] \tag{1.35}$$

This can be rewritten as,

$$L_j(\beta) = \prod_{q=1}^{k_j}\left[\lambda_{0q}\exp\left(\boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]^{\delta_{ijq}} \exp\left[-\lambda_{0q} l_q \exp\left(\boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right] \tag{1.36}$$

The full likelihood is then

$$L(\lambda_{01}, \dots, \lambda_{0r}, \beta) = \prod_{j=1}^{N}\prod_{q=1}^{k_j}\left[\lambda_{0q}\exp\left(\boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]^{\delta_{ijq}} \exp\left[-\lambda_{0q} l_{jq} \exp\left(\boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]$$

$$\propto \prod_{j=1}^{N}\prod_{q=1}^{k_j}\left[\lambda_{0q} l_{jq}\exp\left(\boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]^{\delta_{ijq}} \exp\left[-\lambda_{0q} l_{jq}\exp\left(\boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right] \tag{1.37}$$

The right hand side of (1.37) is a likelihood function of a Poisson model with δ_{ijq} as an outcome. This proportionality can be carried over to a model with random effects. So, the conditional likelihood function of group i is given by

$$L_i(\lambda_{01},\dots,\lambda_{0r},\beta|b_i) = \prod_{j=1}^{n_i}\prod_{q=1}^{k_j}\left[\lambda_{0q}\exp\left(b_i + \boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]^{\delta_{ijq}}\exp\left[-\lambda_{0q}l_{jq}\exp\left(b_i + \boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]$$

$$\propto \prod_{j=1}^{n_i}\prod_{q=1}^{k_j}\left[\lambda_{0q}l_{jq}\exp\left(b_i + \boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]^{\delta_{ijq}}\exp\left[-\lambda_{0q}l_{jq}\exp\left(b_i + \boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right] \quad (1.38)$$

while the marginal likelihood function for all the groups can be written as,

$$L(\lambda_{01},\dots,\lambda_{0r},\beta,\sigma^2) \propto$$

$$\prod_{i=1}^{G}\int_{-\infty}^{\infty}\prod_{j=1}^{n_i}\prod_{q=1}^{k_j}\left[\lambda_{0q}l_{ijq}\exp\left(b_i + \boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]^{\delta_{ijq}}\exp\left[-\lambda_{0q}l_{ijq}\exp\left(b_i + \boldsymbol{\beta}'\boldsymbol{x_{ij}}\right)\right]f(b_i)db_i \quad (1.39)$$

When $b_i \sim N(0,\sigma^2)$, the integral in (1.39) will not be available in a closed form, and integral approximation such as Gauss-Hermite quadrature or Laplace approximation are needed. Instead of estimating a large number of λ_{0q}; $q = 1,\dots,r$, the baseline hazard is modeled as a smooth function of time, e.g., by a 4th order polynomial function (Rabe-Hesketh and Skrondal, 2012). Using available software for GLMMs, nested frailty and random slope Cox models can be estimated.

1.10.4. Bayesian approach

Clayton (1991; 1994) formulated the shared model using the counting-process notation and discussed estimation of the baseline hazard and model parameters using Markov Chain Monte Carlo (MCMC) techniques. The approach is nonparametric with respect to Λ_0, where the cumulative baseline hazard is specified in terms of increments over particular intervals without knowing any information about the hazard function itself. These increments are assumed to be independent and to follow a gamma process. By dividing the follow-up time into intervals with the boundaries corresponding to observed event times as in the Poisson modeling approach, model (1.19) can be estimated by using MCMC methods. The likelihood function for the whole dataset takes the form defined in (1.39), and the joint posterior distribution is given by

$$L(\lambda_{01}, \dots, \lambda_{0K}, \beta, b, \sigma^2)$$

$$\propto L(\lambda_{01}, \dots, \lambda_{0K}, \beta, \sigma^2) \times \pi(b|\sigma^2) \times \pi(\lambda_{01}, \dots, \lambda_{0K}) \times \pi(\beta) \times \pi(\sigma^2) \qquad (1.40)$$

where $\pi(.)$ indicates prior distribution.

Finally, the specification of model extensions to incorporate several hierarchical levels or special structures is straightforward (e.g. Manda, 2001; Yamaguchi et al., 2002).

1.10.5. Penalized partial likelihood approach with Laplace approximation

The approach was proposed by Ripatti and Palmgren (2000) and is one of the more popular approaches for estimating Cox models with normally distributed random effects. Assuming a normal distribution for random effects, the marginal log-likelihood is approximated by the Laplace method for integral approximation. We discuss here the

estimation of the model defined in (1.21) with correlated random effects, the conditional log-likelihood can be written as

$$l_i^C(\lambda_0(.),\beta|b_{i0},b_{i1}) = \sum_{j=1}^{n_i}\left[\lambda_0(y_{ij})\exp(b_{i0} + \beta x_{ij} + b_{i1}x_{ij})\right]^{\delta_{ij}}$$

$$\times \exp\left[-\Lambda_0(y_{ij})\exp(b_{i0} + \beta x_{ij} + b_{i1}x_{ij})\right] \quad (1.41)$$

With $(b_{i0}, b_{i1}) \sim N(0, \Sigma)$, we have

$$f(b_{i0}, b_{i1}) = \frac{1}{2\pi|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(b_{i0}, b_{i1})\Sigma^{-1}(b_{i0}, b_{i1})'\right]$$

Assuming conditional independence of subjects within a group and independence between groups, the marginal log-likelihood for the entire data can be written as

$$l(\lambda_0(.),\beta,\Sigma) = \sum_{i=1}^{G}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left[-K_i(b_{i0}, b_{i1})\right]db_{i0}db_{i1} \quad (1.42)$$

where

$$K_i(b_{i0}, b_{i1}) = -\sum_{j=1}^{n_i}\left[\delta_{ij}(\ln\lambda_0(y_{ij}) + b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right]$$

$$- \Lambda_0(y_{ij})\exp(b_{i0} + \beta x_{ij} + b_{i1}x_{ij})$$

$$+ \ln 2\pi + \frac{1}{2}\ln|\Sigma| + (b_{i0}, b_{i1})\Sigma^{-1}(b_{i0}, b_{i1})'$$

Ignoring the constant $\ln 2\pi$, Ripatti and Palmgren (2000) approximated the marginal likelihood in (1.42) using the Laplace approximation as,

$$l(\lambda_0(.),\beta,\Sigma) \approx \sum_{i=1}^{G}\left[-\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\ln\left|K_i''(\tilde{b}_{i0},\tilde{b}_{i1})\right| - K_i(\tilde{b}_{i0},\tilde{b}_{i1})\right] \qquad (1.43)$$

with $(\tilde{b}_{i0},\tilde{b}_{i1}) = \underset{(b_{i0},b_{i1})\in R^2}{argmax} K_i(b_{i0},b_{i1})$ and the second derivative of $K_i(\tilde{b}_{i0},\tilde{b}_{i1})$ is

$$K_i''(\tilde{b}_{i0},\tilde{b}_{i1}) = \left(\frac{\partial^2 K_i(b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\bigg|_{(\tilde{b}_{i0},\tilde{b}_{i1})}\right)$$

$$= \sum_{j=1}^{n_i}\left[x_{ij}^2\Lambda_0(t)\exp(\tilde{b}_{i0} + \beta x_{ij} + \tilde{b}_{i1}x_{ij})\right] - \Sigma^{-1}$$

So (1.43) can be expressed as

$$l(\lambda_0(.),\beta,\Sigma) \approx \sum_{i=1}^{G}\left[-\frac{1}{2}\ln|\Sigma(\sigma^2)| - \frac{1}{2}ln\left|\left(\frac{\partial^2 K_i(b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\bigg|_{(\tilde{b}_{i0},\tilde{b}_{i1})}\right)\right| + l_i^{PEN}(\lambda_0(.),\beta,\Sigma,\tilde{b}_{i0},\tilde{b}_{i1})\right]$$

where

$$l_i^{PEN}(\lambda_0(.),\beta,\Sigma,\tilde{b}_{i0},\tilde{b}_{i1}) = -K_i(\tilde{b}_{i0},\tilde{b}_{i1})$$

$$= \sum_{j=1}^{n_i}\left[\delta_{ij}\left[\ln\lambda_0(y_{ij}) + \tilde{b}_{i0} + \beta x_{ij} + \tilde{b}_{i1}x_{ij}\right]\right.$$

$$\left. - \Lambda_0(y_{ij})\exp(\tilde{b}_{i0} + \beta x_{ij} + \tilde{b}_{i1}x_{ij})\right]$$

$$-\frac{1}{2}(\tilde{b}_{i0},\tilde{b}_{i1})\Sigma^{-1}(\tilde{b}_{i0},\tilde{b}_{i1})' \qquad (1.44)$$

When Σ is known and (b_{i0},b_{i1}) are considered fixed effect parameters, the (1.44) is a penalized Cox full log-likelihood with b_{i0} and b_{i1} as another set of parameters. The penalized full log-likelihood in (1.44) can be converted into a partial log-likelihood as

$$l_i^{PEN}(\lambda_0(.),\beta,\Sigma,b_{i0},b_{i1})$$

$$= \sum_{j=1}^{n_i}\delta_{ij}\left[b_{i0} + \beta x_{ij} + b_{i1}x_{ij} - \ln\sum_{(p,q)\in R(t_{ij})}\exp\left(b_{p0} + \beta x_{pq} + b_{p1}x_{pq}\right)\right] - \frac{1}{2}(\tilde{b}_{i0},\tilde{b}_{i1})\Sigma^{-1}(\tilde{b}_{i0},\tilde{b}_{i1})'$$

$$+ \sum_{j=1}^{n_i}\delta_{ij}\left[\ln(\lambda_0(t)) + \ln\sum_{(p,q)\in R(t_{ij})}\exp\left(b_{p0} + \beta x_{pq} + b_{p1}x_{pq}\right)\right] - \Lambda_0(t)\exp\left(b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right)$$

$$= l_i^{PPL}(\beta,\Sigma,b_{i0},b_{i1}) + g(\lambda_0(t),\beta,b_{i0},b_{i1}) \qquad (1.45)$$

For fixed Σ and considering (b_{i0},b_{i1}) fixed effect parameters, Ripatti and Palmgren (2000) pointed out that the values $\hat{\beta}(\Sigma)$ and $(\hat{b}_{i0}(\Sigma),\hat{b}_{i1}(\Sigma))$ that maximize $l_i^{PPL}(\beta,\Sigma,b_{i0},b_{i1})$ also maximize $l_i^{PEN}(\hat{\lambda}_0(.),\beta,\Sigma,\tilde{b}_{i0},\tilde{b}_{i1})$ with $\hat{\lambda}_0(t)$ is the estimator of discretized baseline hazard while keeping Σ fixed.

Based on $l_i^{PPL}(\beta,\Sigma,b_{i0},b_{i1})$, the estimating equation for $\beta(\Sigma)$ and $(b_{i0}(\Sigma),b_{i1}(\Sigma))$, given Σ, can be derived. When $\hat{\beta}(\Sigma)$ and $(\hat{b}_{i0}(\Sigma),\hat{b}_{i1}(\Sigma))$ are computed, the matrix Σ can be updated by maximizing the following approximate profile log-likelihood,

$$l(\hat{\beta},\hat{b}_{i0},\hat{b}_{i1},\Sigma) \approx \sum_{i=1}^{G}\left[-\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\ln\left|\left(\frac{\partial^2 K_i(b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\Big|_{(\hat{b}_{i0},\hat{b}_{i1})}\right)\right| - \frac{1}{2}(\hat{b}_{i0},\hat{b}_{i1})\Sigma^{-1}(\hat{b}_{i0},\hat{b}_{i1})'\right] \qquad (1.46)$$

Because of a better empirical performance, Ripatti and Palmgren (2000) suggested replacing $\frac{\partial^2 K_i(b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\Big|_{(\hat{b}_{i0},\hat{b}_{i1})}$ by $\frac{\partial^2 l_i^{PPL}(\beta,\Sigma,b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\Big|_{(\hat{b}_{i0},\hat{b}_{i1})}$ in (1.46).

Finally, the approach is implemented in the coxme package of R software (Therneau, 2013). The coxme library is able to handle different model designs with normally distributed random effects including the shared, random slope, and nested random effects models.

1.11. Brief notes on other approaches

In this section, a brief overview is given for approaches that were not covered in previous sections. One approach to estimating frailty models is to use the expectation-maximization (EM) algorithm (Klein, 1992); the algorithm consists of two steps: the E-step and M-step and iterates between them until convergence is achieved, see Duchateau and Janssen (2008) for detailed description. The EM algorithm approach can fit gamma shared frailty models, and the approach was extended by Cortiñas and Burzykowski (2005) to allow for random slope models with a normal distribution. To our knowledge, it does not support the counting-process format necessary for modeling time-varying predictors and effects.

The hierarchical likelihood approach was proposed by Ha et al. (2001) and can handle shared and nested frailty models with gamma and log-normal distributions. It is implemented in the frailtyHL package of R software (Ha et al., 2012) and provides standard errors for variance components, but it does not allow for time-varying predictors and effects.

The Monte Carlo EM method (MCEM) was proposed by Vaida and Xu (2000) and fits Cox models with normally distributed random effects. The approach uses the EM algorithm along with MCMC at the E-step to compute the conditional expectation of

random effects. The MCEM approach is implemented in the phmm library of R and supports random intercept models as well as independent random intercept and random slope models. Time-dependent predictors and coefficients are not allowed.

The Bayesian approach with Laplace approximation was proposed by Ducrocq and Casella (1996) to analyze survival models with random effects on large datasets in the field of animal genetics, and it was extended by Legrand et al. (2005) to allow for two independent random effects terms within the same cluster. The approach approximates the marginal posterior density using the Laplace approximation. The approach is implemented in the Survival Kit (Ducrocq and Sölkner, 1994), which is a package of Fortran programs that can be found at (http://www.boku.ac.at/nuwi/software/sofskit.htm). The approach can handle time-dependent predictors and effects as well as time-dependent frailty, but not in random slope models.

1.12. Focus and objectives of the thesis

This work focuses on modeling hierarchical data in veterinary science when the response variable is time-to-event. The estimation of mixed effects models that take into account the hierarchical survival data structure is still a topic of intensive research. Many models have been suggested and numerous approaches for estimating these models have been developed in the literature. Beyond a shared frailty model, it is not known at this time which of the existing estimation approaches works best for complex hierarchical survival models. In the present work, the performance of several existing estimation procedures for different hierarchical survival models will be evaluated and compared. In many studies such as those in veterinary epidemiology, datasets can both be extensive in scale

and complex in structure. In addition, the number of predictors of interest may be large, and both time-varying predictors and coefficients may be encountered in the analysis. All of these issues, along with the limitation of software complicate the task of making an appropriate inference. In this thesis, two large veterinary datasets with different structures and time-dependent predictors and coefficients are analyzed.

The main objectives of this thesis are twofold: first, to explore existing estimation methods for multi-component frailty Cox models, and evaluate their performance in terms of bias in point estimates and empirical variability. Second, the study will explore the feasibility of a full hierarchical survival analysis for two large datasets with time-dependent predictors and effects. The specific objective for each chapter along with a brief description is as follows.

In Chapter 2, a review of four estimation methods for a Cox model with random herd and treatment effects; comparing their performance, through simulation, based on a real veterinary dataset. The performance of the methods is investigated in terms of the bias of fixed and random effect estimates and their empirical variability. The aim of the comparison was to establish some practical guidelines for the choice of appropriate statistical estimation procedures for modeling 2-level survival data when a random slope is needed.

In Chapter 3, the feasibility of a full hierarchical survival analysis for a large dataset with time-dependent predictors and coefficients using a log-normal nested frailty Cox model approximated by a mixed-effects Poisson model is explored. A nested frailty Cox model was applied to a 3-level hierarchical survival dataset on clinical mastitis from the

Canadian Bovine Mastitis Research Network, and identified risk factors associated with the hazard of clinical mastitis during the cow's lactation. Further, the performance of the approach used is evaluated and compared with the performance of the penalized partial likelihood approach.

In Chapter 4, a cross-classified and multiple membership Cox model was fit to a large observational dataset on calf loss and mortality in beef cattle from Western Canada. The model is fitted to the data as a mixed-effects Poisson model using MCMC techniques. The individual, herd management, and environmental factors associated with the hazard of calf mortality in Western Canada are examined as well as the age period where the hazard of mortality is highest, is estimated. Moreover, the Poisson GLMM with a Bayesian posterior approach is evaluated via a simulation study using data structures similar to the structure of calf mortality data.

In Chapter 5, through simulation, the estimates of the Poisson maximum likelihood approach with adaptive Gaussian quadrature of estimating Cox model with normal random effects were examined against misspecification of the random-effects distribution. The simulations are performed based on three different hierarchical Cox models and two different non-normal distributions for random effects in each model. Some of the factors that might affect the estimation are also investigated.

In Chapter 6, the general conclusions from this thesis are outlined and some topics for future research are discussed.

References

Aalen O. (1980). A model for nonparametric regression analysis of counting processes. *Lecture Notes in Statistics*, **2**: 1-25.

Aalen O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, **7**: 1121-1137.

Aalen O. (1989). A linear regression model for the analysis of life time. *Statistics in Medicine*, **8**: 907-925.

Aalen O. (1992). Modelling heterogeneity in survival analysis by compound Poisson distribution. *Annals of Applied Probability*, **2**: 951-972.

Bennett S. (1983). Covariance analysis of censored survival data. *Biometrics*, **30**: 89-100.

Breslow N. (1974). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, **2**: 273-277.

Browne W, Goldstein H, Rasbash H. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, **1**, 103-124.

Burnley T, Stryhn H, Burnley H, Hammell L. (2010). Randomized clinical field trial of a bacterial kidney disease vaccine in Atlantic salmon, *Salmo salar* L. *Journal of Fish Diseases*, **33**, 545-557.

Christensen J. (1996). The influence of selected litter and herd factors on treatments for lameness in suckling piglets from 35 Danish herds. *Preventive Veterinary Medicine*, **26**: 107-118.

Clayton D. (1988). The analysis of event history data: a review of progress and outstanding problems. *Statistics in Medicine*, **7**: 819-841.

Clayton D. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, **47**: 467-485.

Clayton D. (1994). Bayesian analysis of frailty models. Technical report, Medical Research Council Biostatistics Unit, Cambridge, UK.

Collett D. (1994). *Modeling Survival Data in Medical Research*. London: Chapman & Hall.

Cortiñas J, Burzykowski T. (2005). A version of the EM algorithm for proportional hazard model with random effects. *Biometrical Journal*, **47**: 847-862.

Cortiñas J, Legrand C, Burzykowski T, Janssen P, Ducrocq V, Duchateau L. (2007). Comparison of different estimation procedures for proportional hazards model with random effects. *Computational Statistics & Data Analysis*, **51**: 3913-3930.

Cox D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, **34**: 187-220.

Dekar M, Magoulick D. (2013). Effects of predators on fish and crayfish survival in Intermittent Streams. *Southeastern Naturalist.* **12**: 197-208.

Dohoo I, Martin W, Stryhn H. (2009). *Veterinary Epidemiologic Research*. 2nd edition, Charlottetown, VER - Inc.

Duchateau L, Janssen P, Lindsey P, Legrand C, Nguti R, Sylvester R. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials. *Computational Statistics & Data Analysis*, **40**: 603-620.

Duchateau L, Janssen P. (2008). *The Frailty Model*. New York: Springer.

Ducrocq V, Sölkner J. (1994). The Survival Kit, a Fortran package for the analysis of survival data. The 5th World Cong. Genet. Appl. Livest. Prod., **22**: 51-52. Dep. Anim. Poultry. Sci., University of Guelph, Guelph, Canada.

Ducrocq V, Casella G. (1996). A Bayesian analysis of mixed survival models. *Genetics Selection Evolution*, **28**: 505-529.

Dufour S, Barkema H, DecCoteaux L, DeVries T, Dohoo I, Reyher K, Roy J, Scholl D. (2010). Development and validation of a bilingual questionnaire for measuring udder health related management practices on dairy farms. *Preventive Veterinary Medicine*, **95**: 74-85.

Duncan C, Stephen C, Campbell J. (2006). Clinical characteristics and predictors of mortality for *Cryptococcus gattii* infection in dogs and cats of southwestern British Columbia. *The Canadian Veterinary Journal*, **47**: 993-998.

Edmondson A, George L, Farver T. (1989). Survival analysis for evaluation of corneal healing times in calves with naturally acquired infectious bovine keratoconjunctivitis. *American Journal of Veterinary Research*, **50**: 838-844.

Efron B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, **72**: 557-565.

Feng S, Wolfe R, Port F. (2005). Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *Journal of the American Statistical Association*, **100**: 718-735.

Fielding A, Goldstein H. (2006). Cross-classified and multiple membership structures in multilevel models: An introduction and review (Research Report Number 791). Birmingham, UK: University of Birmingham, Department of Education and Skills.

Gamst A, Donohue M, Xu R. (2009). Asymptotic properties and empirical evaluation of the NPMLE in the proportional hazards mixed-effects model. *Statistica Sinica*, **19**: 997-1011.

Glidden D, Vittinghoff E. (2004). Modeling clustered survival data from multicentre clinical trials. *Statistics in Medicine*, **23**: 369-388.

Goethals K, Janssen P, Duchateau L. (2008). Frailty models and Copulas: Similarities and differences. *Journal of Applied Statistics*, **35**: 1071-1079.

Grambsch P, Therneau T. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**: 515-526.

Gröhn Y, Gonzalez R, Wilson D, Hertl J, Bennett G, Schulte H, Schukken Y. (2005). Effect of pathogen-specific clinical mastitis on herd life in two New York State dairy herds. *Preventive Veterinary Medicine*, **71**: 105-125.

Ha D, Lee Y, Song K. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, **88**: 233-243.

Ha D, Lee Y, MacKenzie G. (2007). Model selection for multi-component frailty models. *Statistics in Medicine*, **26**: 4790-4807.

Ha I, Noh M, Lee Y. (2012). FrailtyHL: A package for fitting frailty models with H-likelihood. *The R Journal*, **4**: 28-37.

Hangagal D. (2011). *Modelling Survival Data Using Frailty Models*. Chapman & Hall/CRC.

Hangagal D, Dabade A. (2013). A comparative study of shared frailty models for kidney infection data with generalized exponential baseline distribution. *Journal of Data Science*, **11**: 109-142.

Hatcher S, Atkins K, Safari E. (2010). Lamb survival in Australian Merino sheep: A genetic analysis. *Journal of Animal Science*, **88**: 3198-3205.

Heinze G, Schemper M. (2001). A solution to the problem of monotone likelihood in Cox regression. *Biometrics*, **57**: 114-119.

Hirsch K, Wienke A. (2012). Software for semiparametric shared gamma and log-normal frailty models: An overview. *Computer Methods and Programs in Biomedicine*, **107**: 582-597.

Hougaard P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, **1**: 255-273.

Hougaard P. (2000). *Analysis of Multivariate Survival Data*. New-York: Springer Verlag.

Huffler F, McKeague I. (1991). Weighted least squares estimation for Aalen's additive risk model. *Journal of the American Statistical Association*, **86**: 114-129.

Joly P, Commenges D, Letenneur L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics*, **54**: 185-194.

Josiassen M, Christensen J. (1999). Change in lameness risk estimates in piglets due to the modelling of herd-level variation. *Preventive Veterinary Medicine*, **42**: 57-65.

Kaplan E, Meier P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**: 457-481.

Klein J. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics*, **48**: 795-806.

Legrand C, Ducrocq V, Janssen P, Sylvester R, Duchateau L. (2005). A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model. *Statistics in Medicine*, **24**: 3789-3804.

Legrand C. (2010). *Assessing Heterogeneity in Clinical Trials Using the Frailty model*. VDM Verlag Dr. Muller Aktiengesellschaft & Co. KG.

Li D. (2000). On default correlation: a copula function approach. *Journal of Fixed Income*, **9**: 43-54.

Ma R, Krewski D, Burnett R. (2003). Random effects Cox models: A Poisson modelling approach. *Biometrika*, **90**: 157-169.

Manda S. (2001). A comparison of methods for analysing a nested frailty model to child survival in Malawi. *Australian and New Zealand Journal of Statistics*, **43**: 7-16.

Marquardt D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, **11**: 431-441.

Massonnet G, Janssen P, Duchateau L. (2009). Modeling udder infection data using copula models for quadruples. *Journal of Statistical Planning and Inference*, **139**: 3865-3877.

McCorquodale C, Sewalem A, Miglior F, Kelton D, Robinson A, Koeck A, Leslie K. (2013). Short communication: Analysis of health and survival in a population of Ontario Holstein heifer calves. *Journal of Dairy Science*, **96**: 1880-1885.

McGilchrist C, Aisbett C. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**: 461-466.

McGilchrist C. (1993). REML estimation for survival models with frailty. *Biometrics*, **49**: 221-225.

Meadows C, Rajala-Schultz P, Frazer G, Meiring R, Hoblet K. (2006). Evaluation of a contract breeding management program in selected Ohio dairy herds with event-time analysis: I. Cox proportional hazards models. *Preventive Veterinary Medicine*, **77**: 145-160.

Meadows C, Rajala-Schultz P, Frazer G, Phillips G, Meiring R, Hoblet K. (2007). Evaluation of a contract breeding management program in selected Ohio dairy herds with event-time analysis: II. Parametric frailty models. *Preventive Veterinary Medicine*, **80**: 89-102.

Nelsen R. (2006). *An Introduction to Copulas*. Springer, New York.

Nielsen L, Dohoo I. (2012). Survival analysis of factors affecting incidence risk of *Salmonella dublin* in Danish dairy herds during a 7-year surveillance period. *Preventive Veterinary Medicine*, **107**: 160-169.

Nielsen L, Dohoo I. (2013). Time-to-event analysis of predictors from recovery from *Salmonella dublin* infection in Danish dairy herd between 2002 and 2012. *Preventive Veterinary Medicine*, **110**: 370-378.

Nur N, Holmes A, Geupel G. (2004). Use of survival time analysis to analyze nesting success in birds: an example using loggerhead shrikes. *The Condor*, **106**: 457-471.

Portolano B, Finocchiaro R, Van Kaam JB, Riggio V, Maizon D. (2007). Time-to-event analysis of mastitis at first-lactation in Valle del Belice ewes. *Livestock Science*, **110**: 273-279.

Rabe-Hesketh S, Skrondal A. (2012). *Multilevel and Longitudinal Modeling using Stata*. *Volume II: Categorical Responses, Counts, and Survival*, 3rd ed, Stata Press.

Ramsay J. (1988). Monotone regression splines in action. *Statistical Science*, **3**: 425-462.

Relun A, Lehebel A, Bruggink M, Bareille N, Guatteo R. (2013). Estimation of the relative impact of treatment and herd management practices on prevention of digital dermatitis in French dairy herds. *Preventive Veterinary Medicine*, **110**: 558-562.

Reyher K, Dufour S, Barkema H, DecCoteaux L, DeVries T, Dohoo I, Keefe G, Roy J, Scholl D. (2011). The national cohort of dairy farms—A data collection platform for mastitis research in Canada. *Journal of Dairy Science*, **94**: 1616-1626.

Ripatti S, Palmgren J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56**: 1016-1022.

Rondeau V, Gonzalez J. (2003). Maximum penalized likelihood estimation in a Gamma-frailty model. *Lifetime Data Analysis*, **9**: 139-153.

Rondeau V, Filleul L, Joly P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine*, **25**: 4036-4052.

Rondeau V, Michiels S, Liquet B, Pignon J. (2008). Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine*, **27**: 1894-1910.

Rondeau V, Mazroui Y, Gonzalez J. (2012). frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, Volume 47, Issue 4.

Sastry N. (1997). A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of the American Statistical Association*, **92**: 426-435.

Schukken Y, Bar D, Hertl J, Gröhn Y. (2010). Correlated time to event data: Modelling repeated clinical mastitis data from dairy cattle in New York State. *Preventive Veterinary Medicine*, **97**: 150-156.

Shin J, Lu S. (2007). Analysis of failure time data with multilevel clustering with application to the child vitamin A intervention trial in Nepal. *Biometrics*, **63**: 673-680.

Stegeman J, Vernooij J, Khalifa O, Van den Broek J, Mevius D. (2006). Establishing the change in antibiotic resistance of *Enterococcus faecium* strains isolated from Dutch broilers by logistic regression and survival analysis. *Preventive Veterinary Medicine*, **74**: 56-66.

Stryhn H, Christensen J. (2014). The analysis – hierarchical models: past, present and future. *Preventive Veterinary Medicine*, **113**: 304-312.

Tarres J, Tibau J, Piedrafita J, Fabrega E, Reixach J. (2006). Factors affecting longevity in maternal Duroc swine lines. *Livestock Science*, **100**: 121-131.

Therneau T, Grambsch P. (2000). *Modelling Survival Data, Extending the Cox model*. New York: Springer.

Therneau T, Crowson C. (2013). Using time-dependent covariates and time-dependent coefficients in the Cox model. Mayo clinic.

Therneau T. (2013). 'coxme'. A package for mixed effects Cox models. R package version 2.2-3, URL http://cran.r-project.org/web/packages/coxme/.

Vaida F, Xu R. (2000). Proportional hazards model with random effects. *Statistics in Medicine*, **19**: 3309-3324.

Waldner, C. L. (2008). Western Canada study of animal health effects associated with exposure to emissions from oil and natural gas field facilities. Study design and data collection I. Herd performance records and management. *Archives of Environmental & Occupational Health No 4*, **63**: 167-184.

Whitehead J. (1980). Fitting Cox's regression model to survival data using GLIM. *Journal of Applied Statistics*, **29**: 268-275.

Wienke A. (2010). *Frailty Models in Survival Analysis.* Boca Raton: Taylor & Francis.

Xue X, Brookmeyer R. (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime data analysis*, **2**: 227-289.

Yamaguchi T, Ohashi Y. (1999). Investigating centre effects in a multicentre clinical trial of superficial bladder cancer. *Statistics in Medicine*, **18**: 1961-1971.

Yamaguchi T, Ohashi Y, Matsuyama Y. (2002). Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Statistical Methods in Medical Research*, **11**: 221-236.

Yau K. (2001). Multilevel models for survival analysis with random effects. *Biometrics*, **57**: 96-102.

Chapter 2

Comparison of methods for estimating random coefficient Cox models

The work of this chapter is submitted for publishing as: Elghafghuf A, Stryhn H. (2013). Comparison of methods for estimating random coefficient Cox models. *Mathematical Bioscience*. Submitted.

2.1. Abstract

In many studies in medicine, including clinical trials and epidemiological investigations, data are clustered into groups such as health centres (human medicine) or herds (veterinary medicine). Such data are usually analyzed by hierarchical regression models to account for possible variation between groups. When such variation is large, it is of potential interest to explore whether additionally the effect of a within-group predictor varies between groups. In survival analysis, this may be investigated by including two random effects at the group level in a Cox proportional hazards model. Several estimation methods have been proposed to estimate Cox models with additive random effects. We review four of these methods, apply them to real data from veterinary medicine, and compare them using a simulation study.

2.2. Introduction

Survival data from epidemiological veterinary studies involving animals from multiple herds is a typical example of multilevel survival data, also referred to as correlated or clustered survival data. Cox proportional hazards models with random effects within exponential (frailties) acting multiplicatively on an unspecified baseline hazard, are commonly used for multilevel survival data. The Cox model with shared frailties, in which subjects within the same cluster share the same random cluster effect (frailty), provides an intuitive way to describe and quantify the heterogeneity in outcomes. However, this model has some limitations. For example, in shared frailty models the unobserved effect that is not captured by the covariates is assumed to be the same for all subjects within the cluster. Further, the shared frailty model can only induce positive

association within the cluster, despite the possible existence of negative associations (Xue and Brookmeyer, 1996; Wienke et al., 2005). Detailed explanations and excellent examples of the limitations of the shared frailty model are provided by Xue and Brookmeyer (1996).

To address the limitations of the shared frailty model, a Cox model with two or more additive random effects at the cluster level has been studied for the analysis of multilevel survival time data (Xue and Brookmeyer, 1996; Ripatti and Palmgren, 2000; Duchateau and Janssen, 2008; Wienke, 2010). This additive random effects model for example allows for the effect of a treatment at individual level to vary between clusters (i.e. a random coefficient). However estimating Cox models with two correlated additive random effects can be challenging and requires sophisticated techniques to deal with both the unspecified parameter (baseline hazard) and the complex integrals in the likelihood function.

Several estimation procedures have been proposed for estimating model parameters in a Cox model with two additive random effects. For example, Yamaguchi and Ohashi (1999) extended the REML estimation procedure (McGilchrist, 1993), and Ripatti and Palmgren (2000) proposed estimation using a penalized partial likelihood based on Laplace approximation of the marginal likelihood. Furthermore, Vaida and Xu (2000) suggested a Monte Carlo EM algorithm with MCMC sampling technique applied in the E-step, and Ma et al. (2003) reformulated the random effects Cox model as a random effects Poisson model. Cortiñas and Burzykowski (2005) applied a Laplace approximation to the EM algorithm and Legrand et al. (2005) used it to approximate the marginal posterior density in Bayesian approach. Finally, Rondeau et al. (2008)

suggested the use of splines to model the baseline hazard and a Laplace approximation to approximate the marginal likelihood. Massonnet et al. (2008), on the other hand, reformulated the problem of fitting a random coefficient Cox model into a problem of fitting a linear mixed model. Using the estimated integrals of the weighted conditional cumulative log hazard as a linear response, Massonnet et al. (2008) used available software for generalized linear mixed models (GLMM) in estimating model parameters.

Despite the previous work on Cox models with additive random effects, limited work has been done to compare statistical procedures for parameter estimation. A study by Cortiñas et al. (2007) compared different estimation methods based on a Cox model with independent random effects. However, assuming independence between the random effects may lead to invalid assumptions on the variation across clusters (Rondeau et al., 2008). Further, the dependency between the random effects may affect the performance of the estimation methods. In this chapter, we compare four methods of estimating additive random effects Cox models commonly used in epidemiology that are accessible in standard statistical software. These include: the penalized partial likelihood (Ripatti and Palmgren, 2000); the penalized full likelihood (Rondeau et al., 2008); the Poisson maximum likelihood (Rabe-Hesketh and Skrondal, 2012); Bayesian approach (Yamaguchi et al., 2002). For simplicity, we will denote throughout this chapter by PPL, PFL, PML, and BAY the penalized partial likelihood, the penalized full likelihood, the Poisson maximum likelihood, and Bayesian procedures, respectively. Two additive random effects Cox models, one with two independent random effects, and one with two correlated random effects are applied to real data from veterinary science using aforementioned estimation procedures. Through a simulation study, the performance of

these estimation procedures is compared in terms of the bias of fixed and random effect estimates and their empirical variability. This comparison aims at establishing some practical guidelines for the choice of appropriate statistical estimation procedure for modeling treatment variation in 2-level survival data.

2.3. Notation

In the following, we consider clustered survival data from a total of N animals that come from H different herds. The animal j in herd i is either observed from time zero to an event time T_{ij} or to a right censoring time C_{ij} independent of T_{ij}. Let $Y_{ij} = \min(T_{ij}, C_{ij})$ be the observed time and $\delta_{ij} = I_{\{T_{ij} \leq C_{ij}\}}$ be the event indicator. For each animal, we also observe the explanatory variable (predictor) x_{ij}. The simplest model for such data that takes into account the correlation occurring in the data due to clustering is the model with random cluster effects. This model is given by

$$\lambda_{ij}(t|b_{i0}) = \lambda_0(t)\exp(b_{i0} + \beta x_{ij}) \tag{2.1}$$

where $\lambda_{ij}(t|.)$ is the conditional hazard function for the j^{th} animal from the i^{th} herd at time t, $\lambda_0(t)$ is an unspecified baseline hazard at time t, β is a fixed effect parameter, and b_{i0} is the random effect for the i^{th} herd. The random effects $b_{i0}; i = 1, \dots, H$, are assumed to be independent and identically distributed. An alternative formulation of model (2.1) is given by

$$\lambda_{ij}(t|u_{i0}) = \lambda_0(t)u_{i0}\exp(\beta x_{ij}) \tag{2.2}$$

Model (2.2) is known as a shared frailty model, where the frailty $u_{i0} = e^{b_{i0}}$ acts multiplicatively on the baseline hazard. Common choices for the distribution of u_{i0} are one-parameter gamma and log-normal distributions. When u_{i0} are log-normally distributed, the random effects $b_{i0} = \ln(u_{i0})$ follow a normal distribution, and the variance σ^2 of b_{i0} (or u_{i0}) indicates the amount of variation between herds.

When variation between herds exists and is large, a further step is to investigate whether there is variation in the predictor effect between different herds. To do so, an extra random effect is added to model (2.1); this random effect represents an interaction between observable and unobservable (not captured by observed predictors) variables. The Cox model with two additive random effects can be expressed as,

$$\lambda_{ij}(t|b_{i0}, b_{i1}) = \lambda_0(t)\exp\left(b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right) \tag{2.3}$$

where b_{i1} represents the random predictor effect, also termed a random coefficient or random interaction. The random effects b_{i0} and b_{i1} are assumed to be jointly distributed with density function $f(b_{i0}, b_{i1}) \sim N(0, \Sigma)$ with $\Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}$.

Given the random effects (b_{i0}, b_{i1}), observations within herd i are assumed to be independent. Therefore, the conditional likelihood function for herd i is:

$$L_i^C(\lambda_0(.), \beta|b_{i0}, b_{i1}) = \prod_{j=1}^{n_i}\left[\lambda_{ij}\left(y_{ij}|b_{i0}, b_{i1}\right)\right]^{\delta_{ij}} S_{ij}\left(y_{ij}|b_{i0}, b_{i1}\right) \tag{2.4}$$

where

$$S_{ij}(t|b_{i0}, b_{i1}) = \exp\left[-\Lambda_0\left(y_{ij}|b_{i0}, b_{i1}\right)\exp\left(b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right)\right],$$

with $\Lambda_0(t) = \int_0^{y_{ij}} \lambda_0(v)\, dv$ and $(y_{ij}, \delta_{ij}, x_{ij})$ being the observed data for animal j from herd i.

Assuming conditional independence of observations within a herd and independence between herds, the overall marginal likelihood function can be written as,

$$L(\lambda_0(.), \beta, \Sigma) = \prod_{i=1}^{H} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left[-K_i(b_{i0}, b_{i1})\right] db_{i0} db_{i1} \qquad (2.5)$$

where

$$K_i(b_{i0}, b_{i1}) = -ln\left[L_i^C(\lambda_0, \beta | b_{i0}, b_{i1})\right] - \ln[f(b_{i0}, b_{i1})]$$

$$= -\sum_{j=1}^{n_i} \left[\delta_{ij}(\ln \lambda_0(y_{ij}) + b_{i0} + \beta x_{ij} + b_{i1} x_{ij}\right]$$

$$-\Lambda_0(y_{ij}) \exp(b_{i0} + \beta x_{ij} + b_{i1} x_{ij}) - \ln[f(b_{i0}, b_{i1})]$$

The marginal log-likelihood in (2.5) cannot be used directly to estimate the parameters of model (2.3) because it contains an unspecified parameter (λ_0) and depends on integrations that cannot be solved analytically. Several parameter estimation procedures have been proposed to overcome these two problems; four of these parameter estimation procedures are reviewed in next section.

2.4. Parameter estimation procedures

2.4.1. Penalized Partial Likelihood (PPL) procedure

This estimation procedure was proposed by Ripatti and Palmgren (2000) and became in the recent years one of the most commonly used estimation procedures for Cox models with random effects. Based on the derivation of a penalized likelihood solution of Breslow and Clayton (1993) for the GLMM with normal random effects, Ripatti and Palmgren (2000) applied Laplace's method for integral approximation to approximate the marginal log-likelihood by

$$l(\lambda_0(.),\beta,\Sigma)$$

$$\approx \sum_{i=1}^{H}\left[-\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\ln\left|\left(\frac{\partial^2 K_i(b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\Big|_{(\tilde{b}_{i0},\tilde{b}_{i1})}\right)\right| + l_i^{PEN}\left(\lambda_0(.),\beta,\Sigma,\tilde{b}_{i0},\tilde{b}_{i1}\right)\right] \quad (2.6)$$

where $(\tilde{b}_{i0},\tilde{b}_{i1}) = \underset{(b_{i0},b_{i1})\in R^2}{argmax} K_i(b_{i0},b_{i1})$ and

$$l_i^{PEN}\left(\lambda_0(.),\beta,\Sigma,\tilde{b}_{i0},\tilde{b}_{i1}\right) = -K_i\left(\tilde{b}_{i0},\tilde{b}_{i1}\right)$$

$$= \sum_{j=1}^{n_i}\left[\delta_{ij}\left[\ln\lambda_0(y_{ij}) + \tilde{b}_{i0} + \beta x_{ij} + \tilde{b}_{i1}x_{ij}\right] - \Lambda_0(y_{ij})\exp\left(\tilde{b}_{i0} + \beta x_{ij} + \tilde{b}_{i1}x_{ij}\right)\right]$$

$$-\frac{1}{2}(\tilde{b}_{i0},\tilde{b}_{i1})\Sigma^{-1}(\tilde{b}_{i0},\tilde{b}_{i1})'$$

When Σ is known and (b_{i0},b_{i1}) are considered fixed effect parameters, the $l_i^{PEN}(\lambda_0(.),\beta,\Sigma,b_{i0},b_{i1})$ function is a penalized Cox full log-likelihood with b_{i0} and

b_{i1} as another set of parameters. The $l_i^{PEN}(\lambda_0(.),\beta,\Sigma,b_{i0},b_{i1})$ function can be converted into a penalized partial log-likelihood as

$$l_i^{PEN}(\lambda_0(.),\beta,\Sigma,b_{i0},b_{i1})$$

$$= \sum_{j=1}^{n_i}\delta_{ij}\left[b_{i0} + \beta x_{ij} + b_{i1}x_{ij} - \ln \sum_{(p,q)\in R(t_{ij})} \exp\left(b_{p0} + \beta x_{pq} + b_{p1}x_{pq}\right)\right] - \frac{1}{2}(\tilde{b}_{i0},\tilde{b}_{i1})\Sigma^{-1}(\tilde{b}_{i0},\tilde{b}_{i1})'$$

$$+ \sum_{j=1}^{n_i}\delta_{ij}\left[\ln\left(\lambda_0(y_{ij})\right) + \ln \sum_{(p,q)\in R(t_{ij})} \exp\left(b_{p0} + \beta x_{pq} + b_{p1}x_{pq}\right)\right] - \Lambda_0(y_{ij})\exp\left(b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right)$$

$$= l_i^{PPL}(\beta,\Sigma,b_{i0},b_{i1}) + g\left(\lambda_0(y_{ij}),\beta,b_{i0},b_{i1}\right)$$

where $R(t_{ij})$ are risk sets.

For fixed Σ and considering (b_{i0},b_{i1}) fixed effect parameters, Ripatti and Palmgren (2000) pointed out that the values $\hat{\beta}(\Sigma)$ and $(\hat{b}_{i0}(\Sigma),\hat{b}_{i1}(\Sigma))$ that maximize $l_i^{PPL}(\beta,\Sigma,b_{i0},b_{i1})$ also maximize $l_i^{PEN}(\hat{\lambda}_0,\beta,\Sigma,\tilde{b}_{i0},\tilde{b}_{i1})$ with $\hat{\lambda}_0$ is the estimator of discretized baseline hazard while keeping Σ fixed.

Based on $l_i^{PPL}(\beta,\Sigma,b_{i0},b_{i1})$, the estimating equation for $\beta(\Sigma)$ and $(b_{i0}(\Sigma),b_{i1}(\Sigma))$, given Σ, can be derived. When $\hat{\beta}(\Sigma)$ and $(\hat{b}_{i0}(\Sigma),\hat{b}_{i1}(\Sigma))$ are computed, the matrix Σ can be updated by maximizing the following approximate profile log-likelihood,

$$l(\hat{\beta},\hat{b}_{i0},\hat{b}_{i1},\Sigma) \approx \sum_{i=1}^{H}\left[-\frac{1}{2}\ln|\Sigma| - \frac{1}{2}\ln\left|\left(\frac{\partial^2 K_i(b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\bigg|_{(\hat{b}_{i0},\hat{b}_{i1})}\right)\right| - \frac{1}{2}(\hat{b}_{i0},\hat{b}_{i1})\Sigma^{-1}(\hat{b}_{i0},\hat{b}_{i1})'\right] \quad (2.7)$$

Ripatti and Palmgren (2000) suggested replacing $\frac{\partial^2 K_i(b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\Big|_{(\hat{b}_{i0},\hat{b}_{i1})}$ by $\frac{\partial^2 l_i^{PPL}(\beta,\Sigma,b_{i0},b_{i1})}{\partial b_{i0}\partial b_{i1}}\Big|_{(\hat{b}_{i0},\hat{b}_{i1})}$ in (2.7) because it showed better empirical performance in the simulations.

To estimate the covariance matrix of the fixed effect estimates, one can use Cox model software with the estimated random effects as an offset. For estimating the standard error of the estimates of Σ Ripatti and Palmgren (2000) suggested to differentiate (2.7) twice with respect to Σ, and take the expectation with respect to (b_{i0}, b_{i1}). We leave formulas and technical details to the original article.

Finally, the procedure is implemented in the coxme package for R software developed by Therneau (2011) to fit Cox models with normal random effects. The standard error for estimated variances of random effects is not provided in the current implementation.

2.4.2. Penalized Full Likelihood (PFL) procedure

This procedure was proposed by Rondeau et al. (2008); they used the likelihood defined in (2.5) instead of a partial likelihood and penalized the hazard function instead of penalizing the frailties. Rondeau et al. (2008) proposed a smooth baseline hazard estimator and added a penalty term for the roughness of the baseline hazard to the marginal log-likelihood. This roughness penalty term is a product of a smoothing parameter v and the integral of the squared second derivative of the baseline hazard. The penalized log-likelihood is thus defined as:

$$l_{PFL}(\lambda_0(.),\beta,\Sigma) = l(\lambda_0(.),\beta,\Sigma) - v\int_0^{\infty}[\lambda_0''(t)]^2 dt \qquad (2.8)$$

where $\lambda_0''(t)$ is the second derivative of the baseline hazard, and v is a positive smoothing parameter that controls the trade-off between the data fit and the smoothness of the function $\lambda_0(.)$.

For modeling the baseline hazard, Rondeau et al. (2008) suggested to model $\lambda_0(.)$ through splines. As they used cubic M-splines (Ramsay, 1988) that are easy to integrate and differentiate, the second derivative of λ_0 is approximated by a linear combination of 1st order polynomial terms. Such an approximation reduces the number of parameters but still allows for flexible shapes of hazard functions. The approximation error can be reduced by increasing the number of knots. In other words, the more knots are used; the closer is the approximation to the true hazard. The smoothing parameter can be fixed by the user or estimated by maximizing a likelihood cross-validation criterion for the Cox model (Joly et al., 1998; Rondeau and Gonzalez, 2003).

The penalized log-likelihood in (2.8) is maximized by the robust Marquardt algorithm (Marquardt, 1963). This algorithm has the advantage of being stable and fast in convergence. After convergence, the estimated covariance matrix for model parameters is obtained directly from the inverse of converged Hessian matrix.

This procedure is implemented in the frailtypack package for R software (Rondeau et al., 2012). The number of knots for the approximation of the baseline hazard can be controlled by the user and must be between 4 and 20. The smoothing parameter can be automatically estimated by the cross-validation procedure.

2.4.3. Poisson Maximum Likelihood (PML) procedure

One way to fit model (2.3) is to reformulate the Cox model with random effects in a random effects Poisson model framework (Rabe-Hesketh and Skrondal, 2012), as follows. The follow-up period is divided into as many intervals (say K intervals) as there are unique failure times. Each interval begins at a unique failure time and ends at the next unique failure time. This allows estimation of the baseline hazard $\lambda_0(t)$ nonparametrically. For each of these intervals:

$$\lambda_0(t) = \lambda_{0k},\ t \in \Omega_k = (t_{k-1}, t_k],\ k = 1, \dots, K.$$

Let l_{ijk} be the follow-up time of animal j within herd i in Ω_k, and δ_{ijk} be the event indicator for animal j within herd i in Ω_k. As shown in Ma et al. (2003) and Feng et al. (2005), under an independent and non-informative censoring assumption for the interval Ω_k, the conditional likelihood function from a random effects Cox model is proportional to the conditional likelihood function from a random effects Poisson model with log interval lengths between unique failure times as an offset. For the k^{th} interval, the contribution of animal j from herd i to the conditional likelihood is:

$$\begin{aligned} L_{ijk}(\lambda_{0k}, \beta | b_{i0}, b_{i1}) &= \left[\lambda_{0k}\exp\left(b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right)\right]^{\delta_{ijk}} \exp\left[-\lambda_{0k} l_{ijk}\exp\left(b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right)\right] \\ &\propto \left[\lambda_{0k} l_{ijk}\exp\left(b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right)\right]^{\delta_{ijk}} \exp\left[-\lambda_{0k} l_{ijk}\exp\left(b_{i0} + \beta x_{ij} + b_{i1}x_{ij}\right)\right], \end{aligned}$$

corresponding to the Poisson model described above, and the conditional likelihood function of herd i then takes the form,

$$L_i(\lambda_{01}, \dots, \lambda_{0K}, \beta | b_{i0}, b_{i1}) = \prod_{j=1}^{n_i} \prod_{k=1}^{K} L_{ijk}(\lambda_{0k}, \beta | b_{i0}, b_{i1})$$

$$\propto \prod_{k=1}^{K} \prod_{j=1}^{n_i} \left[\lambda_{0k} l_{ijk} \exp\left(b_{i0} + \beta x_{ij} + b_{i1} x_{ij}\right)\right]^{\delta_{ijk}} \exp\left[-\lambda_{0k} l_{ijk} \exp\left(b_{i0} + \beta x_{ij} + b_{i1} x_{ij}\right)\right]$$

while the marginal likelihood function for all herds can be written as,

$$L(\lambda_{01}, \dots, \lambda_{0K}, \beta, \Sigma) \propto$$

$$\prod_{i=1}^{H} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{k=1}^{K} \prod_{j=1}^{n_i} \left[\lambda_{0k} l_{ijk} \exp\left(b_{i0} + \beta x_{ij} + b_{i1} x_{ij}\right)\right]^{\delta_{ijk}} \exp\left[-\lambda_{0k} l_{ijk} \exp\left(b_{i0} + \beta x_{ij} + b_{i1} x_{ij}\right)\right] f(b_{i0}, b_{i1}) db_{i0} db_{i1} \right] \quad (2.9)$$

As in the likelihood function of the previous methods, the integral in (2.9) cannot be solved in a closed form and thus must be approximated. Adaptive Gaussian quadrature can be used for this task; see Pinheiro and Bates (1995) for the details of implementation. Instead of estimating all the parameters $\lambda_{01}, \dots, \lambda_{0K}$, it is customary to model the baseline hazard as a smooth function of time, e.g. by a 4th order polynomial as suggested by Rabe-Hesketh and Skrondal (2012).

2.4.4. Bayesian (BAY) procedure

Bayesian techniques can be used to fit random effects Cox models, where the cumulative baseline hazard is specified in terms of increments over particular intervals without

knowing any information about the hazard function itself. These increments are assumed to be independent and to follow a gamma process. Similar to the idea of fitting a random effects Cox model in a random effects Poisson model described in Subsection 2.4.3, the follow-up time is divided into K intervals with the boundaries corresponding to the observed event times. The likelihood function for the entire data set takes the form defined in (2.9).

The joint density of the posterior distribution for model (2.3) is proportional to

$$L(\lambda_{01}, \dots, \lambda_{0K}, \beta, b, \Sigma) \propto L(\lambda_{01}, \dots, \lambda_{0K}, \beta, \Sigma) \times \pi(b|\Sigma) \times \pi(\lambda_{01}, \dots, \lambda_{0K}) \times \pi(\beta) \times \pi(\Sigma) \quad (2.10)$$

where $b = (b_{i0}, b_{i1})$ and $\pi(.)$ indicates prior distribution.

Conjugate prior distributions are: a normal distribution $N(0, 10^6)$ for the fixed effect parameter β; a gamma distribution $(10^{-3}, 10^{-3})$ for the inverse variances of the two normal random effects if a diagonal covariance matrix is considered for random effects, and a Wishart distribution with a diagonal matrix of 1 and 2 degrees of freedom for the inverse covariance matrix when a more flexible covariance structure for random effects is used. For the baseline hazard, we follow Kalbfleisch (1978) and assume gamma distribution priors for the increments of the baseline hazard with scale $c = 0.001$ and shape parameter equal to $cr\Delta t = 10^{-4}\Delta t$, where $r = 0.1$ is a guess of the failure rate per unit time and Δt is the size of the time interval.

2.4.5. Software implementation for estimation procedures

The PPL and PFL estimation procedures used the previously described implementations in R version 2.14.2 software (the coxme package version 2.1-3 for PPL and the frailtypack package version 2.2 for PFL). A smoothing parameter of 10,000 and 8 knots were used for the PFL estimation procedure. The PML estimation procedure used the adaptive quadrature algorithm for ML estimation implemented in the Stata software version 11.2 with the default number of integration points (7 per random effect). MCMC estimation for the Bayesian model was performed in WinBUGS version 1.4 called from within R software using the R2WinBUGS package (Sturtz et al., 2005).

For Bayesian analyses, we first ran three parallel Markov chains with different initial values for 15,000 iterations and a thinning of 10. All model parameters from the three chains were monitored for convergence. Markov chain diagnostics were carried out for the three chains using the R package coda (Plummer et al., 2006) and found to be satisfactory. The first 5,000 samples after thinning were discarded, and based on further 10,000 samples posterior medians were extracted as model parameter estimates and posterior standard deviations played the role of standard errors.

2.5. Example: Lameness data

The lameness disease dataset originates from a Danish project that was carried out by the Health and Production Surveillance System (HEPS) from October 1990 to March 1991 (Christensen, 1996). The outcome of interest was defined as the (survival) time from birth to the first treatment for lameness (e.g. splayleg, joint infection, or ataxia) in the litter. A total number of 7872 litters of piglets were observed during the period from birth

to weaning. Only litters with a suckling period no more than 40 days were kept, leaving a total of 7632 litters.

In this study, a subset of the data was used in which only one litter per sow was included, and only the 22 herds not participating in any elevated health programs were included. The resulting sample size was 3556 litters of which 398 litters had the event of interest, corresponding to 88.8% censored observations. The number of events in herds ranged from 0 to 69 with a mean of 18 events. The median time of follow-up (till censoring) and median time to event were 27 and 11 days, respectively. The predictor of primary interest here was sow treatment for milk fever, infection, or MMA (mastitis/metritis/agalactia) in days around farrowing (2 days before and up to 4 days after); 26% of the sows were treated. The dataset was analyzed taking into account the variation in the baseline hazard and in the treatment effects between herds using model (2.3). A discussion of the modeling of the full lameness dataset can be found in Stryhn and Christensen (2013).

2.5.1. Analysis of lameness data

A Cox model with a fixed treatment and random treatment by herd interaction was applied to the lameness data using the four estimation procedures reviewed in Section 2.4. To see the impact of including a covariance structure between the random effects, the model was fit to the data with both independent and correlated random effects. The results of the two analyses using the four procedures are shown in Table 2.1. Further, results from a sensitivity analysis based on a guessed failure rate of $r = 0.004$ (estimated based on an exponential model) for the BAY procedure showed very minor changes in model parameter estimates.

The two analyses demonstrated variation in the baseline hazard and in the treatment effect between herds (the log-likelihood values based on PML for a standard Cox model, Cox model with random herd effect, and Cox model with independent random herd and treatment effects were -2531.17, -2331.00 and -2326.97, respectively). This variation was slightly larger in the baseline hazard and a bit smaller in the treatment effect between herds for the independent random effects model than for the correlated random effects model. Furthermore, the variation in the baseline hazard was roughly four times larger than the variation in the treatment effect between herds in the two models. The estimated correlation between random effects in the second analysis ranged between 0.56 (for BAY) and 0.91 (for PML) across the estimation procedures. The estimated values of β from the correlated random effects model were much smaller than those obtained from the independent random effects model even though the correlated random effects model did not show very much improvement in the model fit as the difference in the log-likelihood (or the DIC in Bayesian analysis) between the two models was only approximately one unit (results not shown). As all the estimation procedures agreed on the discrepancy in the fixed effect estimates, one might think of such discrepancy as the result of inadequate assumptions for the variation across herds.

The results for the PFL and PML procedures in both analyses were, in general, quite comparable in terms of the point estimates and slightly different in the standard errors. On the other hand, the results for the PPL and BAY procedures were somewhat different compared with other procedures.

2.6. Simulation studies

The analyses of the real dataset presented in the previous section showed differences in parameter estimates and their standard errors between the four estimation procedures in both independent and correlated random effects models, as well as a strong discrepancy between independent and correlated random effects models. In order to investigate in more detail the performance of the estimation procedures, simulation studies were conducted with settings resembling the settings of the real dataset using the four estimation procedures.

2.6.1. Models and parameter settings

For both the two mixed effects Cox models (with independent or correlated random effects), three different sizes of the heterogeneity parameters were studied and set at: $\sigma_0^2 = 2.0, \sigma_1^2 = 0.5$; $\sigma_0^2 = 0.5, \sigma_1^2 = 0.125$; and $\sigma_0^2 = 0.1, \sigma_1^2 = 0.025$ for large, moderate, and small variance settings, respectively. In each variance setting of the correlated random effects model, three different values of the correlation between the two random effects were considered: $\rho = 0.2$, $\rho = 0.5$, and $\rho = 0.8$. The fixed effect parameter was set at $\beta = -0.4$ in the independent random effects model and at $\beta = -0.8$ in the correlated random effects model. A constant baseline hazard was used in the three simulation settings, and to keep the censoring rate equal in the three settings the baseline hazard was set at 2×10^{-3}, 3.5×10^{-3} and 4.5×10^{-4} for the high, moderate, and low variance settings, respectively.

Finally, the PPL, PFL, and PML procedures were set throughout the simulations as in the analysis of the real dataset. To keep the same covariance matrix for Wishart distribution across the simulation settings of correlated random effects models, the diagonal elements

of the matrix were set at 0.05. This value of the diagonal elements was chosen based on a sensitivity analysis for the values 0.01, 0.05, 0.1, 0.5, and 1. Also, to reduce the computing time of Bayesian analyses in the simulations, MCMC samples with no thinning were used for model estimates.

2.6.2. Simulation of data

To mimic the real dataset, 3556 animals from 22 herds with same sizes (from 68 to 310) as in the real dataset were considered. A total of 300 datasets were generated from model (2.3) using R version 2.14.2 software and the technique of Bender et al. (2005) for generating failure time data. The observations for each particular dataset were generated in the following way: first, the random effects b_{i0} and $b_{i1}; i = 1, \dots, 22$, were generated from a zero-mean bivariate normal distribution $N(0,0, \sigma_0^2, \sigma_1^2, \rho)$. The event time (T_{ij}) for each animal was randomly generated from an exponential distribution with intensity $\lambda_{ij}(t|b_{i0}, b_{i1})$. The same treatment indicator as in the real dataset was used to divide the population into two groups: 26% of the animals in the treatment group and 74% in the control group. The censoring time (C_{ij}) for each animal was randomly generated from the uniform distribution $16 + U(0,24)$. An animal for which the event time T_{ij} was longer than the censoring time C_{ij} was censored with actual time equal to censored time, so that $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = I_{\{T_{ij} \leq C_{ij}\}}$. As in the real data, the amount of censoring was approximately 89% in all the simulation settings.

2.6.3. Analysis of simulated datasets

Model (2.3) was applied to the simulated datasets using the four considered estimation procedures. In order to explore the effect of model misspecification, estimation was carried out for each simulated dataset with and without estimating the covariance between the random herd and treatment effects.

The point estimates of model parameters and their standard errors (posterior standard deviations in Bayesian analysis) for each simulated dataset were extracted. The summary statistics for each model parameter were computed as follows: the mean, computed as the average of the estimated values across the simulated datasets; model-based standard error, computed as the average of the standard errors of the point estimates across the simulated data sets; the empirical standard error, computed as the standard deviation of the estimated values among the simulated datasets; the relative bias, computed as the absolute value of the difference between the mean estimate and the true value divided by the true value; the mean squared error, computed as the mean of the squared differences between the estimated values and the true value over the simulated datasets. If an estimation procedure produced non-sensible estimates or failed to reach convergence for a certain dataset, these results were excluded from the statistics computed across the simulated datasets. Convergence rates across each set of 300 estimations were computed as well.

2.7. Simulation results

2.7.1. Simulated data from independent random effects model

We present first the settings where the true model assumed two independent additive random effects within the same cluster, and the analysis model used either the same

covariance structure (σ_{01} set at 0) as the true model or a more flexible covariance structure where σ_{01} was estimated.

The simulation results for the large, moderate and small variance settings are shown in Table 2.2. The table shows that PFL and PML experienced some convergence difficulties. However, the non-convergence rate was, in general, much higher for PFL than for PML and increased with increasing magnitude of variability of random effects, while the non-convergence rate for PML did not exceed 3% and appeared to be in different directions with the magnitude of random effects variability.

In general, the fixed effect β was estimated well by all the procedures except for PPL in the large variance setting and PFL in the moderate variance setting, where appreciable biases towards zero can be seen. For the PPL method, the bias increased with increasing variance of the random effects as also reported by Ripatti and Palmgren (2000), but this did not seem to be the case for other procedures since no bias pattern was observed. The mean of the estimated SEs and the empirical SEs agreed closely for all the procedures, and PPL had the smallest empirical variability and mean squared error compared with the other three procedures.

For random effects estimates, all procedures produced reasonable estimates for σ_0^2 with exception of PFL in the large variance setting, where a large downwards bias was observed. This can probably be attributed to the very low convergence rate for PFL in this simulation setting. The σ_1^2 was estimated quite well by PML, PPL and PFL in the moderate variance setting and by BAY in the large variance setting. Further, σ_1^2 was overestimated in the small variance setting by all the procedures and underestimated

otherwise. The estimated SEs of the σ_0^2 and σ_1^2 were on average similar to the empirical SEs in all settings for PML, and in the small variance setting for PFL. However, PFL tended to overestimate the SE of σ_0^2 and underestimate it for σ_1^2 in the moderate variance setting. In contrast, the mean of posterior standard deviations for BAY underestimated the empirical SE of σ_0^2 in the moderate variance setting as well as the SEs of both σ_0^2 and σ_1^2 in the large variance setting.

When the analysis allowed the covariance of the random effects to be estimated instead of being set at 0, the simulation results (Table 2.6 in the Appendix) indicated greater convergence difficulties for PFL and PML though the non-convergence rate was still much lower for PML than for PFL. Furthermore, the estimation procedures produced estimates for β and σ_0^2 similar to their estimates in the previous analyses except for BAY, where larger biases for β and σ_0^2 in the moderate variance setting and for σ_0^2 in small variance setting were observed. In the large variance setting and compared with analyses where the covariance was not estimated, σ_1^2 was estimated with less bias by PPL and PML, and with larger bias by BAY, whereas in the small variance setting the four procedure gave estimates with larger biases. For the variability of the estimates, it was noted that the estimated and empirical SEs of β were mostly larger in these analyses than in the analyses where the covariance was not estimated.

2.7.2. Simulated data from correlated random effects model

In this section, the datasets were generated from a model with two correlated additive random effects and analyzed with and without estimating the covariance σ_{01} of the

random effects. The simulation results of the large, moderate and small variance settings with estimation of σ_{01} are shown, respectively, in Tables 2.3, 2.4 and 2.5.

In these simulation settings, only PFL and PML procedures had serious convergence problems, and as in the simulation results of the independent random effects model, the non-convergence rate was much lower for PML than for PFL, and PFL convergence was severely affected by the large variance components.

The estimates for the fixed effect β obtained by the four procedures agreed closely with the true value in the small and moderate variance settings, and their estimated and empirical SEs were close. In contrast, only PML and BAY yielded close estimates for β with nearly unbiased SEs in the large variance setting, whereas PPL underestimated β and its empirical variability. The bias for PPL increased with increasing correlation between the random effects. As in the previous settings, the results for PFL in the large variance setting should be disregarded due to the low convergence rate.

The σ_0^2 was reasonably estimated by all the procedures in the small and moderate variance settings, and by PML and BAY in the large variance setting. In the large variance setting, the PPL procedure produced estimates for σ_0^2 with downwards bias increasing with the correlation between random effects. The estimated and empirical SEs of σ_0^2 agreed well for PML in the three variance settings, and for PFL in the small variance setting. However, the SE of σ_0^2 was overestimated by the BAY approach in all variance settings, and by PFL in the moderate variance setting.

All procedures produced biased estimates for σ_1^2 with exceptions of the PML procedure in the large variance setting and PPL in the moderate variance setting when $\sigma_{01} = 0.05$.

This bias in σ_1^2 estimates was downwards for PPL and BAY in the large variance setting, and for BAY in the moderate variance setting when $\sigma_{01} = 0.05$ and $\sigma_{01} = 0.125$. Otherwise, the bias was upwards. The estimated SEs were close to the empirical SEs for PML and PFL, whereas the mean of the posterior standard deviations for BAY estimator overestimated the empirical SEs.

The four estimation procedures underestimated the covariance between random effects σ_{01} in all cases except in the large variance setting when $\sigma_{01} = 0.5$ and $\sigma_{01} = 0.8$ for PML and BAY, where the procedures yielded estimates somewhat closer to the true values, and when $\sigma_{01} = 0.2$ where the BAY procedure tended to overestimate the σ_{01}. The estimates of SE associated with PML and PFL were closer to the observed SE than the estimates associated with the BAY procedure.

When the analysis ignored the covariance between the random effects, the four procedures showed unsatisfactory results (Tables 2.7, 2.8 and 2.9 in the Appendix). In particular, the procedures tended to produce estimates for the fixed effect β with large biases towards zero when the variance components were large or moderate with strong correlation between the random effects. Furthermore, they yielded estimates for the σ_1^2 with more bias in the large variance setting and less bias in the small variance setting, as well as more bias in the estimates of σ_0^2 for the BAY and less for the PPL in the large variance setting. On the other hand, smaller empirical SEs of β were noted for all the estimation procedures in comparison with the situation where the σ_{01} was estimated as well as smaller estimated SEs for the PFL, PML and BAY, and similar estimated SEs for the PPL.

2.8. Discussion

In this paper, we reviewed four common estimation procedures to fit Cox models with two additive random effects in the same cluster, using data from veterinary epidemiology. We compared these estimation procedures through a simulation study based on the two Cox models with a fixed (treatment) effect and either two independent or two correlated additive random effects. The simulation structure was built to mimic the structure of the real data.

The model used by Cortiñas et al. (2007) is quite similar to the model used in our simulation. Their Cox model had two random effects within the same cluster, but the covariance of the two random effects was set to zero. We used the same model, however we considered a more flexible covariance matrix for the two random effects including their situation where a diagonal covariance matrix was used. What distinguishes our study, besides the flexible structure of the covariance between random effects, is our use of different estimation approaches including the Poisson GLMM approach, with adaptive Gaussian quadrature used for the maximum likelihood estimation. In addition, the settings of the simulation were designed to mimic real data from veterinary medicine, where the magnitude of heterogeneity and censoring rate are often larger than in multicentre trial studies. Finally, our simulations assessed the impact of ignoring the correlation between random effects.

2.8.1. Simulations

2.8.1.1. Convergence and computational requirements for the estimation procedures

Even though neither the PFL nor PML procedures converged in all analyses, the convergence rate was much higher for the latter. PFL exhibited very low convergence rates when the magnitude of heterogeneity was large and even if the magnitude of heterogeneity was relatively small, the optimum convergence rates could not be assured. These low convergence rates remained even after the smoothing parameter was automatically estimated by the cross validation method and a different number of knots was used for the baseline hazard approximation. The non-convergence rate for the PML procedure was much smaller and could be dealt with by changing the integration points for adaptive Gaussian quadrature method.

The PML and BAY procedures were computationally intensive and time consuming (the computing time per dataset were about 10 and 45 minutes for PML and BAY, respectively) because of the need to split the data for PML and the implementation of a gamma process for the BAY procedure. In contrast, PPL and PFL were computationally less intensive (PPL took a few seconds while PFL ran several minutes for one dataset) due to the implementation of Laplace approximation. Furthermore, the PPL procedure was fairly fast to converge because the baseline hazard is estimated simultaneously with other parameters (Feng et al., 2009).

2.8.1.2. Independent versus correlated random effects

When the model was correctly specified, our simulations showed satisfactory results for both independent and correlated random effects models. In case of large variance components, ignoring an existing correlation between random effects resulted in biased estimates for fixed effect coefficients, and this bias increased with increasing σ_{01}. For

instance, when substantial heterogeneity and strong correlation between random effects existed in the data, analyses ignoring the covariance structure led to invalid results. In contrast, taking the covariance structure into account in the analysis when no correlation existed in the data still led to valid estimates. This illustrates the need for a flexible structure for the covariance matrix of random effects. The estimated covariance should be converted to a correlation and assessed relative to its limited range.

2.8.1.3. Estimation of fixed effect parameter

In the presence of limited between-cluster variability for the random effects, all the procedures yielded good and comparable estimates for β. When the magnitude of heterogeneity was large, the picture was somewhat different. Only the PML and BAY procedures gave reasonable estimates for the fixed effect coefficient, whereas it was underestimated by PPL and the PFL estimates were not of interest due to the low convergence rate. Similar findings for PPL were reported in Cortiñas et al. (2007). The estimated and empirical SEs agreed closely for the PML, PFL and BAY procedures. In contrast, PPL underestimated the variability of fixed effect estimates when $\sigma_{01} \neq 0$. This underestimation of fixed effect standard error was pointed out in Ripatti and Palmgren (2000) and Therneau and Grambsch (2000, p. 249).

2.8.1.4. Estimation of random effect parameters

In general, all the procedures performed quite well in estimating σ_0^2, one exception being that the PPL procedure underestimated σ_0^2 when the magnitude of heterogeneity was large. On the other hand, with the exception of PML with large variances, all the estimation procedures tended to produce estimates for σ_1^2 that were mostly biased

upwards if the variance components were small to moderate, and somewhat underestimated for large variances. Cortiñas and Burzykowski (2005) noted the difficulty in estimating variance components for PPL when the magnitude of heterogeneity is relatively small. Similar findings were reported for PFL by Rondeau et al. (2008). The variability of the two variance estimates measured by the empirical standard errors was estimated with reasonable accuracy by PML, and by PFL when the magnitude of heterogeneity was small to moderate, whereas the mean of posterior standard deviation for BAY tended to overestimate it. Finally, the σ_{01} parameter was almost always underestimated by all the estimation procedures. The SE of the BAY estimator for σ_{01} was overestimated, while the SE of the PML and PFL estimators somewhat agreed with the empirical SE.

2.8.1.5. Summary by estimation procedures

The PML procedure performed quite well and converged in most cases. It showed minimal bias for different simulation settings. Nevertheless, it is necessary to keep in mind the size of the expanded data and the required computing time for analyzing such data. The Stata implementation for Poisson GLMMs provided a standard error for the random effects variance, which is an advantage for real applications. PPL was fairly fast to converge and worked fine when the magnitude of the variability of random effects was small to moderate. When the magnitude of the variability of random effects was large, PPL should be used with caution because of the pronounced underestimation of the fixed effect parameter and its SE. Furthermore, the current implementation of PPL in R does not provide SEs for the estimation of the variance components. PFL experienced a lot of convergence failures, even after changing the parameter setting of model specification,

especially in the cases of large variance components. The advantage of PFL is the calculation of the SEs of the random effects variance even though the procedure sometimes produced nonsense values for the SE estimates. In addition to the BAY procedure being time consuming, it resulted in a pronounced underestimation of the variance of treatment random effect. The procedure worked reasonably well for other model parameters, but some posterior standard deviations overestimated the variability between estimates.

2.8.2. Lameness data

The discrepancy in the fixed effect estimates between the independent and correlated random effects models that appeared in the analysis of lameness data was investigated through a simulation study with numerous settings. Generally speaking, findings from the simulations were similar to our findings for the analysis of the lameness data, and thus supported our conclusion that the discrepancy was probably due to model bias from misspecifying the correlation structure. The analysis showed that the effect of treatment across herds varied around an overall "protective" effect ($HR = 0.41$). The results further suggest that herds with higher hazards tended to have a stronger treatment effect, in reflection of the high correlation between the random intercept and random slope. Further exploration of the lameness data could include contextual effects of treatment and time-dependent coefficients for predictor and herd effects (Stryhn and Christensen, 2013), but for simplicity we limited our example to investigating the random treatment effects.

2.8.3. Conclusion

Based on the results of this simulation study, the performance of the considered estimation procedures depends on the magnitude of variability of random effects in the data. The effect of other factors such as the censoring rate, the number of herds, the herd size and the type of explanatory variable on the estimation procedures is beyond the scope of this study and can be a topic for a future research. This study offers practical guidelines for the choice of appropriate statistical procedure for estimating Cox models with two additive random effects. The conclusions can be drawn from the present study are that, (1) the PML procedure appears to be preferable for analysis of clustered survival data with an underlying random effects Cox model; (2) the PPL procedure is suitable for a quick exploration; and (3) estimating the correlation between the two additive random effects in the analysis is always preferable.

Acknowledgements

We thank Dr. Jette Christensen for providing access to the lameness data.

References

Bender R, Augustin T, Blettner M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, **24**: 1712-1723.

Breslow N, Clayton D. (1993). Approximate inference in generalized linear models. *Journal of the American Statistical Association*, **88**: 9-25.

Christensen J. (1996). The influence of selected litter and herd factors on treatments for lameness in suckling piglets from 35 Danish herds. *Preventive Veterinary Medicine*, **26**: 107-118.

Cortiñas J, Burzykowski T. (2005). A version of the EM algorithm for proportional hazard model with random effects. *Biometrical Journal*, **47**: 847-862.

Cortiñas J, Legrand C, Burzykowski T, Janssen P, Ducrocq V, Duchateau L. (2007). Comparison of different estimation procedures for proportional hazards model with random effects. *Computational Statistics & Data Analysis,* **51**: 3913-3930.

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

Feng S, Wolfe R, Port F. (2005). Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *Journal of the American Statistical Association,* **100**: 718-735.

Feng S, Nie L, Wolfe R. (2009). Laplace's approximation for relative risk frailty models. *Lifetime Data Analysis,* **15**: 343-356.

Joly P, Commenges D, Letenneur L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence of dementia. *Biometrics,* **54**: 185-194.

Kalbfleish J. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society: Series B,* **40**: 214-221.

Legrand C, Ducrocq V, Janssen P, Sylvester R, Duchateau L. (2005). A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model. *Statistics in Medicine,* **24**: 3789-3804.

Ma R, Krewski D, Burnett R, (2003). Random effects Cox models: A Poisson modelling approach. *Biometrika,* **90**: 157-169.

Marquardt D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics,* **11**: 431-441.

Massonnet G, Janssen P, Burzykowski T. (2008). Fitting conditional survival models to meta-analytic data by using a transformation toward mixed-effects models. *Biometrics,* **64**: 834-842.

McGilchrist C. (1993). REML estimation for survival models with frailty. *Biometrics,* **49**: 221-225.

Pinheiro J, Bates D. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics,* **4**: 12-35.

Plummer M, Best N, Cowles K, Vines K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News,* **6**: 7-11.

Rabe-Hesketh S, Skrondal A. (2012). *Multilevel and Longitudinal Modeling using Stata. Volume II: Categorical Responses, Counts, and Survival*, 3rd ed, Stata Press.

Ramsay J. (1988). Monotone regression splines in action. *Statistical Science*, **3**: 425-462.

Ripatti S, Palmgren J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56**: 1016-1022.

Rondeau V, Gonzalez J. (2003). Maximum penalized likelihood estimation in a Gamma-frailty model. *Lifetime Data Analysis*, **9**: 139-153.

Rondeau V, Michiels S, Liquet B, Pignon J. (2008). Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine*, **27**: 1894-1910.

Rondeau V, Mazroui Y, Gonzalez J. (2012). frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, Volume 47, Issue 4.

Stryhn H, Christensen J. (2013). The analysis – hierarchical models: past, present and future. *Preventive Veterinary Medicine*, **113**: 304-312.

Sturtz S, Ligges U, Gelman A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, Volume 12, Issue 3.

Therneau T, Grambsch P. (2000). *Modelling Survival Data, Extending the Cox model*. New York: Springer.

Therneau T. (2011). 'coxme'. A package for mixed effects Cox models. R package version 2.1-3, URL http://cran.r-project.org/web/packages/coxme/.

Vaida F, Xu R. (2000). Proportional hazards model with random effects. *Statistics in Medicine*, **19**: 3309-3324.

Wienke A, Arbeev K, Locatelli I, Yashin A. (2005). A comparison of different bivariate correlated frailty models and estimation strategies. *Mathematical Biosciences*, **198**: 1-13.

Wienke A. (2010). *Frailty Models in Survival Analysis.* Boca Raton: Taylor & Francis.

Xue X, Brookmeyer R. (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime data analysis*, **2**: 227-289.

Yamaguchi T, Ohashi Y. (1999). Investigating centre effects in a multicentre clinical trial of superficial bladder cancer. *Statistics in Medicine*, **18**: 1961-1971.

Yamaguchi T, Ohashi Y, Matsuyama Y. (2002). Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Statistical Methods in Medical Research*, **11**: 221-236.

Table 2.1. Parameter estimates (with SE) of independent random effects model (the upper half of the table) and correlated random effects model (the lower half of the table) for the analysis of lameness dataset.

Estimation procedure	β	σ_0^2	σ_1^2	σ_{01}
Independent random effects model				
PPL	-0.353 (0.232)	2.494 (--[a])	0.400 (--[a])	0
PFL	-0.403 (0.280)	2.360 (1.040)	0.412 (0.367)	0
PML	-0.398 (0.274)	2.399 (0.940)	0.401 (0.353)	0
BAY	-0.400 (0.344)	2.663 (0.358)	0.458 (0.376)	0
Correlated random effects model				
PPL	-0.721 (0.227)	2.257 (--[a])	0.492 (--[a])	0.839 (--[a])
PFL	-0.868 (0.493)	2.170 (0.929)	0.564 (0.548)	0.960 (0.765)
PML	-0.891 (0.470)	2.185 (0.868)	0.590 (0.572)	1.036 (0.787)
BAY	-0.788 (0.494)	2.374 (1.147)	0.789 (1.076)	0.771 (0.907)

PPL: Penalized partial likelihood; PFL: Penalized full likelihood; PML: Poisson maximum likelihood; BAY: Bayesian; [a] No available standard error.

Table 2.2. Results of large, moderate, and small variance settings based on correctly specified independent random effects model. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, relative absolute bias, mean squared error and convergence percentage over 300 simulated data sets.

	β			σ_0^2			σ_1^2			σ_{01}	Conv.
Method	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**%**
True	**-.400**			**2.000**			**.500**			**.000**	
PPL	-.351 (.238; .227)	$^{+}$.123	.148	1.874 (.730; ----[a])	$^{-}$.063	.274	.417 (.343; ----[a])	$^{-}$.166	.248	0	100
PFL	-.422 (.256; .263)	$^{-}$.055	.163	1.445 (.416; .587)	$^{-}$.278	.240	.456 (.394; .339)	$^{-}$.088	.312	0	57
PML	-.411 (.253; .252)	$^{-}$.028	.160	1.914 (.748; .717)	$^{-}$.043	.283	.442 (.361; .325)	$^{-}$.116	.266	0	100
BAY	-.417 (.256; .281)	$^{-}$.043	.165	2.122 (.839; .297)	$^{+}$.061	.358	.474 (.433; .297)	$^{-}$.052	.376	0	100
True	**-.400**			**.500**			**.125**			**.000**	
PPL	-.376 (.160; .164)	$^{+}$.060	.065	.506 (.213; ----[a])	$^{+}$.012	.090	.123 (.144; ----[a])	$^{-}$.016	.168	0	100
PFL	-.360 (.166; .175)	$^{+}$.100	.073	.483 (.186; .250)	$^{-}$.034	.070	.120 (.154; .125)	$^{-}$.040	.192	0	78
PML	-.396 (.171; .175)	$^{+}$.010	.073	.478 (.203; .182)	$^{-}$.044	.084	.125 (.154; .129)	$^{+}$.000	.192	0	98
BAY	-.400 (.169; .182)	$^{+}$.000	.073	.519 (.219; .147)	$^{+}$.038	.096	.106 (.149; .199)	$^{-}$.152	.176	0	100
True	**-.400**			**.100**			**.025**			**.000**	
PPL	-.398 (.133; .138)	$^{+}$.005	.045	.101 (.056; ----[a])	$^{+}$.010	.030	.034 (.057; ----[a])	$^{+}$.360	.120	0	100
PFL	-.380 (.137; .142)	$^{+}$.050	.048	.105 (.062; .056)	$^{+}$.050	.040	.032 (.055; .052)	$^{+}$.280	.120	0	99
PML	-.408 (.136; .144)	$^{-}$.020	.048	.092 (.052; .036)	$^{-}$.080	.030	.034 (.059; .042)	$^{+}$.360	.160	0	97
BAY	-.416 (.135; .150)	$^{-}$.040	.045	.096 (.060; .089)	$^{-}$.040	.040	.031 (.040; .146)	$^{+}$.240	.080	0	100

[a] No available standard error; $^{+}$ upwards bias; $^{-}$ downwards bias.

Table 2.3. Results of large variance setting based on correctly specified correlated random effects model. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, relative absolute bias, mean squared error, and convergence percentage over 300 simulated data sets.

	β			σ_0^2			σ_1^2			σ_{01}			Conv.
Method	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**%**
True	**-.800**			**2.000**			**.500**			**.800**			
PPL	-.671 (.347; .216)	$^{+}$.161	.170	1.758 (.632; ----[a])	$^{-}$.121	.228	.419 (.333; ----[a])	$^{-}$.162	.234	.618 (.429; ----[a])	$^{-}$.228	.205	100
PFL	-.918 (.256; .406)	$^{-}$.148	.234	1.334 (.554; .515)	$^{-}$.333	.373	.530 (.452; .472)	$^{+}$.060	.404	.589 (.466; .387)	$^{-}$.264	.323	19
PML	-.794 (.368; .350)	$^{+}$.008	.169	1.878 (.737; .711)	$^{-}$.061	.279	.502 (.390; .405)	$^{+}$.004	.302	.734 (.489; .459)	$^{-}$.083	.303	90
BAY	-.793 (.355; .352)	$^{+}$.009	.156	1.928 (.764; .888)	$^{-}$.036	.293	.469 (.357; .509)	$^{-}$.062	.256	.768 (.483; .552)	$^{-}$.040	.293	100
True	**-.800**			**2.000**			**.500**			**.500**			
PPL	-.689 (.333; .228)	$^{+}$.139	.154	1.806 (.676; ----[a])	$^{-}$.097	.247	.421 (.354; ----[a])	$^{-}$.158	.262	.396 (.425; ----[a])	$^{-}$.208	.456	100
PFL	-.868 (.386; .389)	$^{-}$.085	.189	1.131 (.257; .514)	$^{-}$.435	.410	.518 (.410; .454)	$^{+}$.036	.332	.312 (.422; .376)	$^{-}$.376	.422	24
PML	-.786 (.356; .355)	$^{+}$.018	.159	1.869 (.708; .708)	$^{-}$.066	.259	.477 (.380; .401)	$^{-}$.046	.288	.460 (.479; .452)	$^{-}$.080	.460	97
BAY	-.785 (.351; .344)	$^{+}$.019	.154	1.919 (.743; .886)	$^{-}$.041	.279	.380 (.318; .464)	$^{-}$.240	.230	.511 (.464; .516)	$^{+}$.022	.430	100
True	**-.800**			**2.000**			**.500**			**.200**			
PPL	-.707 (.333; .237)	$^{+}$.116	.149	1.849 (.684; ----[a])	$^{-}$.076	.245	.425 (.354; ----[a])	$^{-}$.150	.260	.155 (.431; ----[a])	$^{-}$.225	1.46	100
PFL	-.857 (.354; .398)	$^{-}$.071	.159	1.191 (.374; .539)	$^{-}$.405	.397	.530 (.404; .474)	$^{+}$.060	.324	.107 (.404; .395)	$^{-}$.465	.850	29
PML	-.783 (.358; .355)	$^{+}$.021	.160	1.886 (.698; .717)	$^{-}$.057	.250	.474 (.376; .396)	$^{-}$.052	.284	.180 (.490; .451)	$^{-}$.100	1.20	99
BAY	-.774 (.353; .340)	$^{+}$.033	.156	1.928 (.735; .894)	$^{-}$.036	.272	.324 (.283; .422)	$^{-}$.352	.222	.232 (.467; .500)	$^{+}$.160	1.09	100

[a] No available standard error; $^{+}$ upwards bias; $^{-}$ downwards bias.

Table 2.4. Results of moderate variance setting based on correctly specified correlated random effects model. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, relative absolute bias, mean squared error, and convergence percentage over 300 simulated data sets.

	β			σ_0^2			σ_1^2			σ_{01}			Conv.
Method	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**%**
True	**-.800**			**.500**			**.125**			**.200**			
PPL	-.759 (.217; .178)	$^{+}$.051	.061	.493 (.198; ----[a])	$^{-}$.014	.078	.145 (.149; ----[a])	$^{+}$.160	.184	.167 (.151; ----[a])	$^{-}$.165	.120	100
PFL	-.812 (.261; .266)	$^{-}$.015	.085	.484 (.195; .268)	$^{-}$.032	.076	.149 (.166; .173)	$^{+}$.192	.224	.151 (.164; .175)	$^{-}$.245	.145	78
PML	-.794 (.238; .227)	$^{+}$.008	.071	.467 (.188; .182)	$^{-}$.066	.072	.146 (.156; .161)	$^{+}$.168^{+}	.200	.167 (.155; .149)	$^{-}$.165	.125	90
BAY	-.794 (.219; .227)	$^{+}$.008	.060	.473 (.194; .219)	$^{-}$.054	.076	.140 (.111; .186)	$^{+}$.120	.096	.163 (.135; .167)	$^{-}$.185	.100	100
True	**-.800**			**.500**			**.125**			**.125**			
PPL	-.766 (.222; .179)	$^{+}$.043	.063	.495 (.200; ----[a])	$^{-}$.010	.080	.136 (.145; ----[a])	$^{+}$.088	.168	.103 (.158; ----[a])	$^{-}$.176	.200	100
PFL	-.786 (.266; .271)	$^{+}$.018	.089	.504 (.219; .291)	$^{+}$.008	.096	.146 (.165; .176)	$^{+}$.168	.216	.082 (.170; .179)	$^{-}$.344	.248	81
PML	-.804 (.244; .228)	$^{-}$.005	.074	.469 (.185; .184)	$^{-}$.062	.070	.142 (.158; .165)	$^{+}$.136	.200	.107 (.163; .151)	$^{-}$.144	.216	95
BAY	-.799 (.225; .222)	$^{+}$.001	.063	.470 (.195; .219)	$^{-}$.060	.078	.116 (.095; .170)	$^{-}$.072	.072	.108 (.139; .162)	$^{-}$.136	.152	100
True	**-.800**			**.500**			**.125**			**.050**			
PPL	-.775 (.223; .180)	$^{+}$.031	.063	.495 (.197; ----[a])	$^{-}$.010	.078	.129 (.148; ----[a])	$^{+}$.032	.176	.034 (.163; ----[a])	$^{-}$.320	.540	100
PFL	-.768 (.269; .260)	$^{+}$.050	.091	.490 (.209; .252)	$^{-}$.020	.088	.150 (.177; .182)	$^{+}$.200	.256	.004 (.184; .170)	$^{-}$.920	.720	78
PML	-.804 (.243; .225)	$^{-}$.005	.074	.467 (.190; .184)	$^{-}$.066	.074	.135 (.164; .166)	$^{+}$.080	.216	.038 (.167; .151)	$^{-}$.240	.560	98
BAY	-.806 (.226; .216)	$^{-}$.008	.064	.467 (.193; .218)	$^{-}$.066	.076	.097 (.080; .154)	$^{-}$.224	.056	.050 (.140; .159)	$^{+}$.000	.380	100

[a] No available standard error; $^{+}$ upwards bias; $^{-}$ downwards bias.

Table 2.5. Results of small variance setting based on correctly specified correlated random effects model. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, relative absolute bias, mean squared error, and convergence percentage over 300 simulated data sets.

	β			σ_0^2			σ_1^2			σ_{01}			Conv.
Method	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**%**
True	**-.800**			**.100**			**.025**			**.040**			
PPL	-.797 (.195; .163)	$^{+}$.004	.048	.102 (.020; ----[a])	$^{+}$.020	.030	.069 (.080; ----[a])	$^{+}$1.76	.320	.023 (.062; ----[a])	$^{-}$.425	.100	100
PFL	-.803 (.218; .192)	$^{-}$.004	.059	.105 (.063; .058)	$^{+}$.050	.040	.076 (.090; .112)	$^{+}$2.04	.480	.024 (.065; .065)	$^{-}$.400	.125	88
PML	-.818 (.208; .183)	$^{-}$.023	.054	.094 (.055; .051)	$^{-}$.060	.030	.072 (.085; .108)	$^{+}$1.88	.360	.023 (.063; .064)	$^{-}$.425	.100	94
BAY	-.817 (.194; .179)	$^{-}$.021	.048	.093 (.050; .058)	$^{-}$.070	.030	.061 (.037; .105)	$^{+}$1.44	.120	.017 (.037; .059)	$^{-}$.575	.050	100
True	**-.800**			**.100**			**.025**			**.025**			
PPL	-.799 (.191; .163)	$^{+}$.001	.045	.102 (.019; ----[a])	$^{+}$.020	.030	.067 (.078; ----[a])	$^{+}$1.68	.320	.011 (.063; ----[a])	$^{-}$.560	.160	100
PFL	-.800 (.212; .190)	.000	.056	.105 (.063; .058)	$^{+}$.050	.040	.073 (.088; .104)	$^{+}$1.92	.400	.014 (.066; .063)	$^{-}$.440	.160	91
PML	-.819 (.209; .182)	$^{-}$.024	.055	.094 (.055; .051)	$^{-}$.060	.030	.074 (.088; .109)	$^{+}$1.96	.400	.011 (.065; .064)	$^{-}$.560	.160	91
BAY	-.821 (.189; .178)	$^{-}$.026	.045	.092 (.050; .058)	$^{-}$.080	.030	.058 (.031; .103)	$^{+}$1.32	.080	.010 (.036; .058)	$^{-}$.600	.080	100
True	**-.800**			**.100**			**.025**			**.010**			
PPL	-.806 (.190; .163)	$^{-}$.008	.045	.102 (.019; ----[a])	$^{+}$.020	.030	.065 (.078; ----[a])	$^{+}$1.60	.320	-.001 (.063; ----[a])	$^{-}$1.10	.400	100
PFL	-.808 (.216; .189)	$^{-}$.018	.059	.105 (.063; .059)	$^{+}$.060	.040	.074 (.091; .108)	$^{+}$1.36	.440	-.000 (.070; .063)	$^{-}$1.20	.500	86
PML	-.825 (.205; .181)	$^{-}$.031	.053	.094 (.055; .051)	$^{-}$.060	.030	.069 (.083; .106)	$^{+}$1.76	.360	.000 (.063; .065)	$^{-}$1.00	.400	94
BAY	-.829 (.189; .177)	$^{-}$.036	.045	.091 (.049; .057)	$^{-}$.090	.020	.055 (.026; .100)	$^{+}$1.20	.080	.003 (.035; .058)	$^{-}$.700	.100	100

[a] No available standard error; $^{+}$ upwards bias; $^{-}$ downwards bias.

Appendix

Table 2.6. Results of large, moderate, and small variance settings based on misspecified independent random effects model. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, relative absolute bias, mean squared error, and convergence percentage over 300 simulated data sets.

	β			σ_0^2			σ_1^2			σ_{01}			Conv.
Method	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**abs.bias**	**mse**	**%**
True	**-.400**			**2.000**			**.500**			**.000**			
PPL	-.357 (.286; .228)	$^{+}$.108	.208	1.879 (.730; ----[a])	$^{-}$.061	.273	.449 (.366; ----[a])	$^{-}$.102	.272	-.005 (.384; ----[a])	$^{-}$.005	.147	100
PFL	-.428 (.298; .352)	$^{-}$.070	.220	1.116 (.276; .511)	$^{-}$.442	.428	.496 (.465; .396)	$^{-}$.008	.426	.039 (.313; .335)	$^{+}$.039	.098	25
PML	-.418 (.320; .320)	$^{-}$.045	.255	1.920 (.749; .728)	$^{-}$.040	.283	.483 (.387; .359)	$^{-}$.034	.298	.010 (.424; .422)	$^{+}$.010	.179	99
BAY	-.399 (.310; .307)	$^{+}$.003	.240	1.964 (.783; .904)	$^{-}$.018	.306	.343 (.331; .397)	$^{-}$.314	.268	.049 (.427; .470)	$^{+}$.049	.184	100
True	**-.400**			**.500**			**.125**			**.000**			
PPL	-.383 (.172; .165)	$^{+}$.043	.075	.505 (.215; ----[a])	$^{+}$.010	.092	.140 (.157; ----[a])	$^{+}$.120	.200	.000 (.135; ----[a])	$^{+}$.000	.018	100
PFL	-.368 (.204; .235)	$^{+}$.080	.105	.485 (.176; .261)	$^{-}$.030	.062	.159 (.185; .160)	$^{+}$.272	.280	-.012 (.153; .157)	$^{-}$.012	.023	67
PML	-.406 (.191; .198)	$^{-}$.015	.090	.476 (.207; .187)	$^{-}$.048	.086	.145 (.168; .149)	$^{+}$.160	.056	.002 (.141; .139)	$^{+}$.002	.020	97
BAY	-.446 (.234; .205)	$^{-}$.115	.143	.598 (.452; .278)	$^{+}$.196	.426	.121 (.153; .164)	$^{-}$.032	.184	.054 (.222; .179)	$^{+}$.054	.052	100
True	**-.400**			**.100**			**.025**			**.000**			
PPL	-.397 (.137; .140)	$^{+}$.008	.048	.104 (.060; ----[a])	$^{+}$.040	.040	.049 (.061; ----[a])	$^{+}$.960	.160	-.009 (.047; ----[a])	$^{-}$.009	.002	100
PFL	-.392 (.146; .153)	$^{+}$.020	.053	.097 (.058; .053)	$^{-}$.030	.030	.051 (.062; .077)	$^{+}$1.04	.200	-.005 (.042; .052)	$^{-}$.005	.002	72
PML	-.401 (.140; .151)	$^{-}$.003	.050	.097 (.058; .053)	$^{-}$.030	.030	.047 (.055; .075)	$^{+}$.880	.160	-.009 (.048; .056)	$^{-}$.009	.002	91
BAY	-.415 (.138; .157)	$^{-}$.038	.048	.146 (.169; .081)	$^{+}$.460	.300	.052 (.029; .083)	$^{+}$1.08	.080	.002 (.051; .065)	$^{-}$.002	.003	100

[a] No available standard error; $^{+}$ upward bias; $^{-}$ downward bias.

Table 2.7. Results of large variance setting based on misspecified correlated random effects model. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, relative absolute bias, mean squared error and convergence percentage over 300 simulated data sets.

	β			σ_0^2			σ_1^2			σ_{01}	Conv.
Method	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**%**
True	**-.800**			**2.000**			**.500**			**.800**	
PPL	-.369 (.280; .216)	$^{+}$.539	.330	1.963 (.739; ----[a])	$^{-}$.019	.273	.300 (.255; ----[a])	$^{-}$.400	.210	0	100
PFL	-.659 (.361; .328)	$^{+}$.176	.185	1.184 (.280; .563)	$^{-}$.408	.372	.393 (.408; .383)	$^{-}$.214	.350	0	26
PML	-.427 (.302; .265)	$^{+}$.466	.288	2.045 (.790; .771)	$^{+}$.023	.312	.330 (.282; .286)	$^{-}$.340	.216	0	98
BAY	-.429 (.311; .296)	$^{+}$.464	.293	2.268 (.893; .309)	$^{+}$.134	.434	.340 (.348; .298)	$^{-}$.320	.292	0	100
True	**-.800**			**2.000**			**.500**			**.500**	
PPL	-.495 (.282; .225)	$^{+}$.381	.215	1.912 (.717; ----[a])	$^{-}$.044	.260	.335 (.272; ----[a])	$^{-}$.330	.202	0	100
PFL	-.701 (.348; .307)	$^{+}$.124	.163	1.191 (.278; .528)	$^{-}$.405	.365	.413 (.335; .380)	$^{-}$.174	.238	0	29
PML	-.551 (.289; .260)	$^{+}$.311	.181	1.846 (.866; .696)	$^{-}$.077	.386	.349 (.303; .299)	$^{-}$.302	.230	0	100
BAY	-.562 (.303; .297)	$^{+}$.298	.185	2.182 (.834; .304)	$^{+}$.091	.363	.386 (.362; .309)	$^{-}$.228	.288	0	100
True	**-.800**			**2.000**			**.500**			**.200**	
PPL	-.625 (.281; .235)	$^{+}$.219	.136	1.891 (.726; ----[a])	$^{-}$.055	.268	.374 (.313; ----[a])	$^{-}$.252	.228	0	100
PFL	-.787 (.344; .304)	$^{+}$.016	.146	1.191 (.270; .528)	$^{-}$.405	.364	.445 (.373; .397)	$^{-}$.110	.282	0	33
PML	-.692 (.299; .272)	$^{+}$.135	.126	1.914 (.709; .721)	$^{-}$.043	.255	.406 (.337; .341)	$^{-}$.188	.244	0	100
BAY	-.696 (.308; .302)	$^{+}$.130	.131	2.124 (.794; .300)	$^{+}$.062	.322	.413 (.387; .319)	$^{-}$.174	.314	0	100

[a] No available standard error; $^{+}$ upward bias; $^{-}$ downward bias.

Table 2.8. Results of moderate variance setting based on misspecified correlated random effects model. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, relative absolute bias, mean squared error, and convergence percentage over 300 simulated data sets.

	β			σ_0^2			σ_1^2			σ_{01}	Conv.
Method	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**%**
True	**-.800**			**.500**			**.125**			**.200**	
PPL	-.650 (.201; .177)	$^{+}$.188	.079	.537 (.214; ----[a])	$^{+}$.074	.094	.120 (.147; ----[a])	$^{-}$.040	.176	0	100
PFL	-.679 (.219; .200)	$^{+}$.151	.078	.505 (.193; .266)	$^{+}$.010	.074	.116 (.162; .146)	$^{-}$.072	.208	0	73
PML	-.677 (.216; .200)	$^{+}$.154	.078	.508 (.204; .195)	$^{+}$.016	.082	.127 (.161; .151)	$^{+}$.016	.208	0	98
BAY	-.676 (.214; .207)	$^{+}$.155	.076	.552 (.225; .153)	$^{+}$.104	.106	.101 (.145; .223)	$^{-}$.192	.168	0	100
True	**-.800**			**.500**			**.125**			**.125**	
PPL	-.695 (.204; .178)	$^{+}$.131	.066	.521 (.208; ----[a])	$^{+}$.042	.088	.118 (.144; ----[a])	$^{-}$.056	.168	0	100
PFL	-.718 (.218; .201)	$^{+}$.103	.068	.501 (.196; .264)	$^{+}$.002	.076	.120 (.159; .152)	$^{-}$.040	.200	0	78
PML	-.721 (.213; .198)	$^{+}$.099	.065	.491 (.195; .189)	$^{-}$.018	.076	.119 (.148; .149)	$^{-}$.048	.176	0	98
BAY	-.721 (.212; .205)	$^{+}$.099	.064	.534 (.215; .151)	$^{+}$.068	.094	.093 (.137; .222)	$^{-}$.256	.160	0	100
True	**-.800**			**.500**			**.125**			**.050**	
PPL	-.744 (.203; .179)	$^{+}$.070	.055	.503 (.198; ----[a])	$^{+}$.006	.078	.108 (.183; ----[a])	$^{-}$.136	.152	0	100
PFL	-.755 (.214; .199)	$^{+}$.056	.060	.489 (.189; .249)	$^{-}$.022	.072	.116 (.159; .149)	$^{-}$.072	.200	0	81
PML	-.769 (.215; .196)	$^{+}$.039	.059	.473 (.189; .182)	$^{-}$.054	.072	.112 (.156; .142)	$^{-}$.104	.192	0	98
BAY	-.773 (.211; .203)	$^{+}$.034	.056	.516 (.207; .149)	$^{+}$.032	.086	.084 (.134; .220)	$^{-}$.328	.160	0	100

[a] No available standard error; $^{+}$ upward bias; $^{-}$ downward bias.

Table 2.9. Results of small variance setting based on misspecified correlated random effects model. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, relative absolute bias, mean squared error, and convergence percentage over 300 simulated data sets.

Method		β			σ_0^2			σ_1^2			σ_{01}	Conv.
		mean (std; se)	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**r.bias**	**mse**	**mean (std; se)**	**%**
	True	**-.800**			**.100**			**.025**			**.040**	
PPL		-.778 (.186; .161)	$^{+}$.028	.044	.106 (.059; ----[a])	$^{+}$.060	.030	.058 (.087; ----[a])	$^{+}$1.32	.360	0	100
PFL		-.756 (.184; .166)	$^{+}$.055	.045	.110 (.065; .058)	$^{+}$.100	.040	.043 (.064; .075)	$^{+}$.720	.160	0	92
PML		-.772 (.179; .167)	$^{+}$.035	.041	.098 (.057; .051)	$^{-}$.020	.040	.043 (.062; .074)	$^{+}$.720	.160	0	87
BAY		-.804 (.191; .178)	$^{-}$.005	.045	.101 (.062; .093)	$^{+}$.010	.040	.045 (.060; .182)	$^{+}$.800	.160	0	100
	True	**-.800**			**.100**			**.025**			**.025**	
PPL		-.787 (.181; .161)	$^{+}$.016	.041	.104 (.057; ----[a])	$^{+}$.040	.030	.053 (.080; ----[a])	$^{+}$1.12	.280	0	100
PFL		-.763 (.179; .166)	$^{+}$.046	.041	.108 (.063; .057)	$^{+}$.080	.040	.043 (.064; .073)	$^{+}$.720	.160	0	93
PML		-.784 (.180; .167)	$^{+}$.020	.040	.096 (.056; .050)	$^{-}$.040	.030	.042 (.064; .073)	$^{+}$.680	.160	0	86
BAY		-.812 (.185; .177)	$^{-}$.015	.043	.098 (.060; .092)	$^{-}$.020	.040	.042 (.053; .178)	$^{+}$.680	.120	0	100
	True	**-.800**			**.100**			**.025**			**.010**	
PPL		-.800 (.181; .160)	$^{+}$.000	.041	.101 (.056; ----[a])	$^{+}$.010	.030	.048 (.076; ----[a])	$^{+}$.920	.240	0	100
PFL		-.781 (.181; .165)	$^{+}$.024	.041	.104 (.061; .056)	$^{+}$.040	.040	.038 (.060; .071)	$^{+}$.520	.160	0	92
PML		-.801 (.180; .166)	$^{-}$.001	.040	.094 (.055; .049)	$^{-}$.060	.030	.036 (.057; .069)	$^{+}$.440	.120	0	89
BAY		-.825 (.185; .177)	$^{-}$.031	.044	.095 (.059; .092)	$^{-}$.050	.030	.038 (.049; .175)	$^{+}$.520	.120	0	100

[a] No available standard error; $^{+}$ upward bias; $^{-}$ downward bias.

Chapter 3

Survival analysis of clinical mastitis data using a nested frailty Cox model fit as a mixed-effects Poisson model

The work of this chapter is submitted for publishing as: Elghafghuf A, Dufour S, Reyher K, Dohoo I, Stryhn H. (2014). Survival analysis of clinical mastitis data using a nested frailty Cox model fit as a mixed-effects Poisson model. *Preventive Veterinary Medicine*, in revision.

3.1. Abstract

Mastitis is a complex disease affecting dairy cows and is considered to be the most costly disease of dairy herds. The hazard of mastitis is a function of many factors, both managerial and environmental, making its control a difficult issue to milk producers. Observational studies of clinical mastitis (CM) often generate datasets with a number of characteristics which influence the analysis of those data: the outcome of interest may be the time to occurrence of a case of mastitis, predictors may change over time (time-dependent predictors), the effects of factors may change over time (time-dependent effects), there are usually multiple hierarchical levels, and datasets may be very large. Analysis of such data often requires expansion of the data into the counting-process format – leading to larger datasets – thus complicating the analysis and requiring excessive computing time.

In this study, a nested frailty Cox model with time-dependent predictors and effects was applied to Canadian Bovine Mastitis Research Network data in which 10,831 lactations of 8,035 cows from 69 herds were followed through lactation until the first occurrence of CM. The model was fit to the data as a Poisson model with nested normally distributed random effects at the cow and herd levels. Risk factors associated with the hazard of CM during the lactation were identified, such as parity, calving season, herd somatic cell score, pasture access, fore-stripping, and proportion of treated cases of CM in a herd. The analysis showed that most of the predictors had a strong effect early in lactation and also demonstrated substantial variation in the baseline hazard among cows and between herds. A small simulation study for a setting similar to the real data was conducted to evaluate the Poisson maximum likelihood estimation approach with both Gaussian quadrature method and Laplace approximation. Further, the performance of the two methods was compared with the performance of a widely used estimation approach for frailty Cox models based on the penalized partial likelihood. The simulation study showed good performance for the Poisson

maximum likelihood approach with Gaussian quadrature and biased variance component estimates for both the Poisson maximum likelihood with Laplace approximation and penalized partial likelihood approaches.

3.2. Introduction

Worldwide, mastitis is one of the most common and costly diseases that affect dairy cattle (Schepers and Dijkhuizen, 1991). Mastitis is a multifactorial disease of the bovine udder and can be caused by multiple bacterial pathogens. Mastitis may be either clinical, if the infection signs are discernible with the naked eye, or subclinical, when no signs are visible and laboratory techniques are needed to detect the infection (Barkema et al., 1998; Olde Riekerink et al., 2008). The hazard of mastitis can be elevated by both environmental factors and managerial practices. Determining what factors or practices might cause CM and then taking necessary action to prevent the disease from occurring is of high priority for dairy producers. Therefore, much research has been dedicated to identifying risk factors associated with CM incidence under different conditions and countries (e.g. Barkema et al., 1999; Barnouin et al., 2005; O'Reilly et al., 2006; Green et al., 2007; Nyman et al., 2007), and many statistical methods have been suggested for analysis of clinical mastitis data (Gasqui and Barnouin, 2003; Schukken et al., 2010). Actually, many complexities inherent to CM data (e.g. censored observations, clustering of observations, recurrence of CM events), impede applicability of most analytical methods, especially in large datasets.

For instance, the Canadian Bovine Mastitis Research Network (CBMRN) launched in 2006 a two-year longitudinal data collection of 91 herds from four regions of Canada. This resulted in the generation of one of the largest, most comprehensive mastitis databases ever assembled, which has supported research into both clinical and subclinical mastitis in areas as

diverse as risk factor evaluation to vaccine development. Details of the data collection platform have been previously published (Reyher et al., 2011).

Survival models have been extensively used in mastitis research over the last 15 years, e.g. for modeling CM event times (Rupp and Boichard, 2000; Santos et al., 2004; Gröhn et al., 2005; Goethals et al., 2009; Schukken et al., 2010; Borne et al., 2011). The most widely used analytical approach is based on Cox's semiparametric proportional hazards model (Cox, 1972; Ducrocq and Casella, 1996). In proportional hazards Cox models, the hazard function is described as a product of unspecified baseline hazard and an exponential function of the multiplied vectors of predictors and regression parameters. One of the strengths of the Cox model is its ability to encompass predictors and coefficients that may vary over time (Therneau and Crowson, 2013). For example, management practice data may be collected at points in time and practices may change between assessments. Dufour et al. (2010), for instance, showed that 27% and 33% of dairy producers reported at least one modification of their milking and housing management procedures, respectively, over a 6-month period. In addition, the effects of explanatory variables often vary over time (e.g. a factor may exert a different effect in early lactation compared to later in lactation). Finally, CM data may be collected from different lactations within the same cow, and with multiple cows located in the same herds or farms. In this scenario, the data are clustered within cows and herds. When clustering is present in the data, the independence assumption of the standard Cox proportional hazards model is no longer valid as event times within the same cluster are correlated. When analyzing such data, it is important to account for the hierarchical structure of the data and take into account the time-dependent predictors and effects.

Extending the Cox proportional hazards model by the addition of random effects (frailty terms) has been proposed for modeling clustered survival data (Therneau and Grambsch,

2000; Duchateau and Janssen, 2008; Wienke, 2010; Hanagal, 2011). In these frailty models, the heterogeneity caused by unobserved factors (due to clustering) are quantified by frailty (or random effects) terms corresponding to the hierarchical levels. In a 3-level hierarchical data structure (e.g. lactations within cows and cows within herds), nested frailty models account for hierarchical clustering by including two nested frailties that act multiplicatively on the baseline hazard (Sastry, 1997; Rondeau et al., 2006). Assuming a log-normal distribution for the nested frailty terms (or normally distributed random effects on the log scale), the nested frailty model can be estimated in a Poisson mixed-effects model framework, after transforming the data to a counting-process format and modeling the baseline hazard as a smooth function of time (Ma et al., 2003; Feng et al., 2005; Rabe-Hesketh and Skrondal, 2012). With the Poisson modeling approach, either adaptive Gaussian quadrature or Laplace approximation can be applied to the likelihood function. However, expanding the dataset in this manner results in a potentially very large dataset which compromises computing time.

The aims of this study were threefold. Firstly, risk factors associated with the hazard of clinical mastitis during the lactation in CBMRN data were evaluated. Secondly, the feasibility of a full hierarchical survival analysis for a large dataset with time-dependent predictors and coefficients using a log-normal nested frailty Cox model approximated by a mixed-effects Poisson model was evaluated. Thirdly, the performance, in terms of bias and efficiency of estimates, of the Poisson maximum likelihood approach (estimated using either Gaussian quadrature or Laplace approximation) was compared with one of the existing estimation approaches, namely, the penalized partial likelihood approach (Ripatti and Palmgren, 2000) through a simulation study. This step served to validate the results obtained for the previous objectives.

3.3. Materials and methods

3.3.1. Data

Data were extracted from the National Cohort of Dairy Farms (NCDF) database, collected from January 2007 to December 2008. The details of the NCDF, including herd selection process and data collection, have been described elsewhere (Reyher et al., 2011). Briefly, a total of 91 dairy herds located in four different geographic regions of Canada participated in the NCDF data collection. Information related to CM, such as the time of occurrence and pathogen species or group of species, were recorded. Dairy Herd Improvement (DHI) data from 2006 to 2008 were used to capture supplementary information on individual and herd levels, such as calving date, culling date, dry-off date, lactation number, herd somatic cell score (SCS), and herd demographics. Moreover, information related to herd management was collected on four different occasions during the study via the udder health related management survey described in Dufour et al. (2010). These four time periods were defined as follows: Jan.-May 2007, Jun.-Dec. 2007, Jan.-May 2008, and Jun.-Dec. 2008. Because CM cases were sampled and recorded by farmers, and to avoid bias resulting from incomplete reporting of CM cases, only the data from NCDF herds showing sufficient compliance for CM sampling were considered in the current study. Compliance was assessed by comparing the number of CM samples submitted for bacteriological analyses and the number of CM cases recorded in the producer's computerized health records. Herds in which the number of submitted CM cases was less than 80% of the number of recorded CM cases were excluded from the analysis. This brought down the number of herds to 74, and an additional 5 herds were excluded due to incomplete records, leaving 69 herds and 265 herd-periods as the information for some of these herds were not available for all 4 periods. Lactations that

started before the onset of the study (i.e. before January 2007) were not included in the analysis.

Within the lactation, event (clinical mastitis) time was defined as the number of days from calving until developing a first case of CM. Observations from cows that did not experience the event within the lactation and were either culled, dried off, left the cohort, or followed-up until the end of the study period were considered right censored. A list of explanatory variables in the dataset is presented in Table 3.1.

3.3.2. Statistical model

A Cox model with two nested frailties to account for the data structure present in the CBMRN data was considered. The nested frailty Cox model (Rondeau et al., 2006) can be written as,

$$\lambda_{ijk}(t|u_i, u_{ij}) = \lambda_0(t)u_i u_{ij} \exp\left[\boldsymbol{\beta}'(t)\boldsymbol{X}_{ijk}(t)\right] \quad (3.1)$$

where $\lambda_{ijk}(t|u_i, u_{ij})$ is a conditional hazard function for lactation k of cow j in herd i conditional on the two nested frailties u_i and u_{ij}; $\lambda_0(t)$ is an unspecified baseline hazard function, $\boldsymbol{X}_{ijk}(t)$ is a vector of (possibly time-varying) predictors with corresponding parameter vector (possibly time-varying) effects $\boldsymbol{\beta}(t)$. An alternative formulation for model (3.1) is given by

$$\lambda_{ijk}(t|b_i, b_{ij}) = \lambda_0(t) \exp\left[b_i + b_{ij} + \boldsymbol{\beta}'(t)\boldsymbol{X}_{ijk}(t)\right] \quad (3.2)$$

where $b_i = \log u_i$ and $b_{ij} = \log u_{ij}$ are nested random effects, and assumed to be normally distributed with zero means and variances σ_c^2 and σ_h^2 for cow and herd levels, respectively.

As described in Rabe-Hesketh and Skrondal (2012), the Cox regression model can be estimated by a Poisson regression model with a specific offset and binary response, and estimated using Poisson maximum likelihood (PML). This approximation can be carried over to the Cox model with random effects (Ma et al., 2003; Feng et al., 2005; Feng et al., 2009; Rabe-Hesketh and Skrondal, 2012). Using available software for generalized linear mixed models (GLMMs), for example the xtmepoisson command in Stata, model (3.2) can be fit to multilevel survival data as a mixed-effects Poisson model after expanding the data to the counting process format and using the length of each time interval on the logarithm scale as an offset. The baseline hazard can be modeled as a polynomial function of time, in this case a fourth-order polynomial function as suggested in Rabe-Hesketh and Skrondal (2012). Finally, for a Poisson maximum likelihood estimation, the Laplace approximation or adaptive Gaussian quadrature can be applied.

3.3.3. Statistical analysis

3.3.3.1. Descriptive statistics and unconditional associations

Descriptive analyses were carried out for each variable in the dataset individually to determine distributions and detect unlikely observations; in this step, time-varying predictors were identified. Correlations or associations among all explanatory variables were computed to assess for possible collinearity. Unconditional associations (simple) between explanatory variables and the hazard of CM were estimated in standard Cox regression models while accounting for time-varying predictors as needed. The proportional hazards assumption was evaluated for every predictor by a statistical test based on the scaled Schoenfeld residuals (Dohoo et al., 2009). The functional form of continuous variables was evaluated by a lowess smoothing graph of the continuous variable against the martingale residuals. A liberal p-value of 0.20 was chosen to determine potentially important explanatory variables.

3.3.3.2. Model building towards a final model

A stepwise backward selection strategy including significant predictors from the unconditional analyses was used to construct a multivariable model for the hazard of CM based on model (2) using the PML approach. In the selection process, a liberal p-value of 0.10 was used as variable inclusion criterion, and potential confounders (such as region) were kept in the model regardless of their significance. After eliminating variables with non-significant effects on the hazard of CM during the model building process, possible and biologically meaningful interactions between explanatory variables in the model were evaluated, and significant interactions were kept. Next, the proportional hazards assumptions for each predictor, individually as well as globally, were assessed as described above in a non-hierarchical multivariable Cox model. For predictors showing non-constant hazards, suitable interaction terms with time were included in the model and assessed using the PML approach. The assumption of independent censoring was checked by a sensitivity analysis in the non-hierarchical Cox model, as described by Dohoo et al. (2009).

3.3.3.3. Software and global settings

The descriptive statistics, unconditional and final model analyses were performed in Stata/MP 12.1. Due to the excessive computing time required for each analysis, all model building analyses were submitted as jobs to the Atlantic Computational Excellence Network (ACEnet), a cluster of computers in a Linux environment platform (http://www.ace-net.ca). In the model building process, maximum likelihood estimation based on the Laplace approximation implemented in the lme4 library of R software (version 2.13.1) was applied. In the final model analysis, adaptive Gaussian quadrature with 8 integration points was used for more accurate likelihood approximation, and inference was based on a significance level of

0.05. The computing time for the final model (dataset consisted of 1,947,560 rows after expansion) exceeded 48 hours on a PC with a 2nd generation Intel® Core™ i3 processor.

3.4. Simulation study

A simulation study was conducted to evaluate the performance of the PML approach discussed above for estimating log-normal nested frailty Cox models. The performance of the PML approach with both adaptive Gaussian quadrature (PMLGQ) and Laplace approximation (PMLAP) for likelihood estimation was compared with the performance of the commonly used approach based on penalized partial likelihood (PPL) proposed by Ripatti and Palmgren (2000). The PPL approach uses Laplace approximation for the marginal likelihood, and is implemented in the coxme package in R (Therneau, 2013). The simulation study was designed with settings resembling the settings of the CBMRN data.

3.4.1. Setup of the simulation study

The simulated datasets were generated from the nested random effects proportional hazards model defined in (3.2). The model included herd and cow random effects and 3 explanatory variables: a trichotomous lactation-level predictor and dichotomous predictors at both cow and herd levels. The 3 predictors and their effects were assumed constant over time as the implementation of PPL does not support the counting-process data setup that is required to handle time varying predictors.

A total of 250 simulated datasets were generated in R software using the technique of Bender et al. (2005). The hierarchical structure and the cluster and subcluster sizes in each simulated dataset matched the hierarchical structure and the cluster and subcluster sizes of the CBMRN dataset. Each simulated dataset was generated as follows: the herd and cow random effects were generated independently from normal distributions with mean zero and variances σ_h^2

and σ_c^2 for herd and cow random effects, respectively. The fixed-effects variables at cow and herd levels were fixed across simulations and randomly generated from Bernoulli distributions with probability of 0.5, while parity in the real data (1: 1st lactation, 2: 2nd lactation, 3: ≥3rd lactation) was used as the lactation-level variable. The event time for each lactation (T_{ijk}) was randomly generated from the model defined in (3.2) based on a Weibull baseline hazard with scale and shape parameters of 0.0035 and 0.62, respectively, as preliminary analysis using a Weibull model showed a strongly significant shape parameter. The censoring time for each lactation, C_{ijk}, was randomly generated from a mixture distribution with proportions of 55% from a lognormal distribution with mean of 4.2 and variance of 0.45, and 45% from a uniform distribution on $[1, 220]$. The censoring time was truncated at 713 days to reflect the length of censored time intervals in the real data. A lactation for which the event time T_{ijk} was longer than the censored time C_{ijk} was censored with actual time equal to censoring time, so the actual time $Y_{ijk} = \min(T_{ijk}, C_{ijk})$. The event indicator δ_{ijk} was generated to be equal to 1 if the event time was shorter than the censored time and was set to 0 otherwise. The censoring rate in this simulation setting matched the censoring rate of the real data.

The summary statistics for each model parameter in the simulation were computed as follows: the estimate was computed as the mean of the estimated values across the simulated datasets; model standard errors were computed as the average of the standard errors of the estimates across the simulated datasets; the empirical standard deviation was computed as the standard deviation of the estimated values among the simulated datasets; the relative bias was computed as the absolute value of the difference between the mean estimate and the true value divided by the true value. Finally, the statistics computed across the simulated datasets

were based only on the analyses when sensible results were produced and convergence was achieved.

3.5. Results

3.5.1. Analysis of CBMRN data

3.5.1.1. Descriptive analysis

A total of 10,831 lactations of 8,035 cows from 69 herds were used in the analysis; 5,264 cows represented a single lactation, and 2,746 and 25 cows represented 2 and 3 lactations, respectively. The herd size ranged from 33 to 345 cows, and median herd CM incidence rate observed in the 69 herds was 21.3 per 100 cow–year (1 year = 305 days) days with 25^{th} and 75^{th} percentiles of 12.3 and 27.9, respectively. The final model included 1,536 CM cases (i.e. 86% censored observations). The medians of event and censoring times were 75 and 245 days, respectively. Kaplan-Meier survival curves for CM events for cows of parity 1, 2 and 3+ are displayed in Figure 3.1. The distributions of each level of the categorical variables and descriptive statistics of the continuous variables for lactation- and herd-level predictors are shown in Tables 3.2 and 3.3, respectively.

3.5.1.2. Multivariable model analysis

The final model included parity (1, 2, and 3+), calving season (1, 2, 3 and 4), mean of herd somatic cell score, fore-stripping (0, 1), pasture access (0, 1), proportion of CM cases treated with an antimicrobial ($\leq 50\%$, $> 50\%$), and geographic region (1, 2, 3, and 4). The final model also included interactions between parity and pasture access, and between parity and proportion of CM cases treated.

Preliminary analysis showed the highest hazard of CM early in lactation. This induced strongly non-proportional hazards for many predictors. Because of this, and to ease the interpretation, predictors were allowed to have different effects in two time periods within the lactation: before and after 13 days in milk (DIM). The cut-point of 13 DIM was chosen because it gave a better model fit compared with other cut-points in the range 10-50 DIM. For those predictors that showed a changing hazard over time, an interaction term between the predictor in question and time (modeled as a binary variable representing the two time periods in the lactation) was added to the model. These interaction terms included a three-way interaction as the interaction between parity and pasture access showed non-constant hazards over time. Results of the final model are presented in Table 3.4. The effect of each predictor is discussed below in turn.

Fore-stripping

The use of fore-stripping in herds was borderline significant ($p = 0.059$) as a predictor for the hazard of CM. The hazard ratio (HR) for cows in herds that used fore-stripping, relative to cows from herds that did not use it, was estimated at 1.41 [95% CI; 0.99; 1.98]. Therefore, the hazard of CM at any given time during the lactation was 41% higher for cows in herds that used fore-stripping.

Calving season

The effect of calving season on the hazard of CM was different early and later in the lactation. In the first 13 DIM, at any given day the highest hazard of CM was for cows that calved in spring and summer compared with both autumn and winter. For instance, the HRs during the first 13 DIM for cows that calved in spring and summer relative to winter were computed, respectively, to be $\exp(-.043 + .687) = 1.90$ [$p = 0.007; 95\% = 1.33 - 2.74$] and 2.48 [$p < 0.001; 95\% = 1.74 - 3.53$]. No significant difference between autumn and

winter was observed during the early lactating period. On the other hand, after 13 DIM the hazard was significantly lower in autumn than in winter ($HR = \exp(-0.327) = 0.72$; $p < 0.001$; $95\% = 0.60 - 0.87$) and also than in spring ($HR = 0.75$; $p = 0.035$; $95\% = 0.62 - 0.91$), but there were no significant differences among spring, summer and winter calving after 13 DIM.

Herd somatic cell score

There was no significant effect of herd SCS after 13 DIM. In the first 13 DIM, the hazard of CM was negatively associated with the herd SCS ($p = 0.017$). For example, the HRs at 10% (SCS of 2.1), 25% (2.4), and 75% (3.0) percentiles of the distribution of herd SCS relative to the mean (2.7) were estimated, respectively, to be 1.26 ($95\% = 1.04 - 1.51$), 1.13 ($95\% = 1.02 - 1.25$), and 0.88 ($95\% = 0.79 - 0.98$) indicating that the hazard of CM at any given time in the first 13 DIM was greatest for herds that had a low mean of SCS.

Parity, pasture access, and cases of CM treated with antimicrobials

The estimated coefficients for all the combinations of cow parity, pasture access, and proportion of cases of CM treated in the first and after 13 DIM are presented in Figure 3.2, and corresponding HRs and their 95% confidence intervals are tabulated in Table 3.5. The effect of pasture access on the hazard of CM was different in the two time intervals of the lactation and depended on cow parity. Furthermore, the coefficients of the interaction between pasture access and cow parity were also time-dependent.

In general, hazard of CM tended to increase with increasing parity. For all combinations of pasture access, DIM, and proportion of CM treated, $\geq 3^{rd}$ lactation cows always showed higher hazard of CM than 2^{nd} lactation cows. These differences, however, were not significant during the > 13 DIM period for cows with no pasture access. Similarly, 1^{st}

lactation cows generally showed lower hazard of CM than 2nd or ≥ 3rd lactation cows. This was not the case, however, during the first 13 DIM for 1st lactation cows that did not have access to pasture. In this later situation, 1st parity cows showed significantly higher hazard of CM than 2nd lactation cows. Furthermore, these cows showed hazard of CM fairly similar to ≥ 3rd lactation cows (Figures 3.1 and 3.2).

In the first 13 DIM, pasture access reduced the hazard of CM in primiparous cows, but it was statistically non-significant in 2nd and ≥ 3rd lactation cows. The HR for primiparous cows having access to pasture between 0-13 DIM was estimated at 0.36 [$p = .002; 95\% = 0.19 - 0.68$] when compared with primiparous cows with no pasture access during that period. In contrast, the effect of pasture access on the hazard of CM was statistically non-significant after 13 DIM irrespective of the cow's parity.

The analysis showed that the proportion of cases of CM treated in the herd had different effects on the hazard of CM early and later in the lactation. Such effects were relatively weak and statistically non-significant early in the lactation. Although the effect of proportion of cases treated appeared to be slightly higher after 13 DIM, this later effect was only close to significant in biparous and multiparous cows. After 13 DIM, a high proportion of cases of CM treated in a herd appeared to be associated with an elevated hazard of CM in biparous and multiparous cows. For instance, after 13 DIM, the estimated HR for biparous cows from herds with a high proportion of treated cases of CM was 1.44 [$p = .093; 95\% = 0.94 - 2.20$].

Non-significant associations

The variable region had no significant effect on the hazard of CM, but it was included in the multivariable model because it showed a strong confounding effect for the association

between mean herd SCS and CM hazard (based on relative difference between unconditional and conditional parameter estimates; difference of 41%).

3.5.2. Simulation study

The simulation results along with the true values of the fixed effects and variance component parameters are reported in Table 3.6. The PPL estimation procedure converged in all simulation iterations, whereas the PMLGQ and PMLAP procedures failed to reach convergence for 2% and 28% of the datasets, respectively. The convergence difficulties for PMLGQ can be solved by adjusting the number of integration points of the adaptive Gaussian quadrature method. There was a substantial difference in the computing time between PPL and both PMLGQ and PMLAP. While PPL took a few minutes to run on a PC, PMLGQ and PMLAP needed several hours to run for one dataset.

The lactation-level (β_1 and β_2 representing parity), cow-level (β_3), and herd-level fixed effects (β_4) were estimated well by all the estimation procedures with relative biases not more than 5.3%, 5%, and 2% for PPL, PMLAP, and PMLGQ, respectively. The confidence interval (CI) coverage for the β_2-β_4 estimates of PPL and PMLGQ, and the β_3 estimate of PMLAP were close to nominal ($\geq$ 94% coverage), whereas the CI coverage for the β_1 estimates of PPL and PMLGQ showed a slight CI under-coverage (92 and 93%, respectively). Furthermore, PMLAP produced estimates for β_1, β_2, and β_4 with important CI under-coverage (ranging from 84 to 91%). The model-based standard error and the empirical variability of fixed effects estimates agreed closely for PPL and PMLGQ estimates while PMLAP underestimated the variability of fixed effects coefficients.

The PMLGQ procedure estimated the between-cow and between-herd variances with relative biases of less than 4% and 2%, respectively, while PMLAP strongly overestimated the between-cow and underestimated the between-herd variances. The PPL estimates for the

between-cow variance were strongly downward biased (−32%); on the other hand, the between-herd variance was estimated well, with relative bias less than 5%. The model-based standard errors of variance components produced by PMLGQ and the empirical standard deviations were fairly close, whereas PMLAP strongly underestimated these standard errors. The PPL procedure produced variance component estimates with the smallest empirical variability but its current implementation in R software does not provide standard errors for these estimates.

3.6. Discussion

3.6.1. Analysis of CBMRN data

3.6.1.1. Incidence rates

At first sight the median first CM incidence rate observed in the current study (21.3 cases/100 cow-y) appeared to be fairly similar to what have been reported in a previous Canadian study (23.0 cases/100 cow-y; Olde Riekerink et al., 2008). In that study, however, second and third cases of clinical mastitis were also included. When considering the recurrent nature of the disease, the current study's first CM incidence rate is, therefore, probably substantially higher than the unreported first CM case incidence rate of the Olde Riekerink et al. (2008) study. One common problem with clinical mastitis research is the often important underreporting of CM events by dairy producers. For instance, second or third CM events in a given lactation, less severe cases, or CM events on quarters for which a persistent infection has already been identified will often go unreported (Vaarst et al., 2002). Although these CM events are of interest from the researcher's perspective, most dairy producers will not see any practical benefit in reporting these, often resulting in incomplete records. In the current study, the selection of herds showing a certain level of compliance regarding CM reporting (based on

reporting consistency) may have resulted in more complete recording and is probably responsible for the relatively higher CM incidence rate.

The higher CM hazard observed during the early lactating period in the current study is very similar to observations made in previous studies conducted in Canada (Olde Riekerink et al., 2008), the USA (Pinedo et al., 2012), France (Gasqui et al., 2003), the Netherlands (Steeneveld et al., 2008) and the UK (Green et al., 2002). In a study by Green et al. (2002), > 50% of the infections resulting in CM cases occurring in the first 30 DIM were deemed to have been acquired during the dry-off period. In comparison 80% of CM cases occurring later during the lactation were the result of new infections acquired during the lactation (Green et al., 2002). Given the drastically different management between dry-off and lactating periods and the different etiology of infections acquired during these two periods, different effects for a given risk factor on the risk of early *vs.* later lactation CM is to be expected. This was the case in the current study, as the measure of effect of all predictors, except fore-stripping cows as part of the milking routine, were significantly modified by DIM. After investigating different thresholds including 10, 12, 13, 14, 15, 25, and 50 DIM, the cut-point of 13 DIM was the best based on the model fit (the best model fit was defined as the model with the greatest likelihood). This division of the lactation time acted as an approximation for the effects of time on the original scale in order to reduce noise~~s~~ and simplify model interpretation since the final model incorporated complex interaction effects.

3.6.1.2. Fore-stripping

The positive association between fore-stripping cows as part of the milking procedures and hazard of CM has been reported in several studies (Elbers et al., 1998; O'Reilly et al., 2006; Richert et al., 2013). This association is very likely a case of reverse causation. Fore-stripping is essential to uncover mild cases of CM; these mild cases often go unnoticed in herds where

the milkers do not check the foremilk at milking time, hence the observed higher hazard of CM in herds using fore-stripping. The hazard difference associated with this practice would be more meaningfully interpreted as an increased risk of CM detection. In the current study, for instance, it may be hypothesized that fore-stripping increased the risk of detecting CM by 40% (i.e. HR: 1.4; 95% CI: 0.99, 2.0). Furthermore, this association was only borderline significant (*P*-value: 0.059).

3.6.1.3. Calving season

Calving season was significantly associated with CM hazard. Seasonality of CM hazard has often been observed in countries having temperate climate. In most studies, higher CM hazard has been observed in cows calving during warmer seasons. For instance, higher CM incidence was observed for cows calving between June-September in Pennsylvania (Erskine et al., 1988) and Wisconsin, USA (Pantoja et al., 2009). Similarly, a higher risk of CM during the first month of lactation was observed for cows calving between June-November in the Netherlands (Steeneveld et al., 2008). In Norway, higher incidence of CM was observed in first parity cows calving between April-August (Waage et al., 1998). The same trend for greater hazard of CM during the early lactating period for cows calving during warmer months was observed in the current study. We can hypothesize that the combined effect of the higher counts of environmental bacteria in the bedding or immediate environment of the cow usually seen in warmer months and of the increased stress and immunosuppression associated with hot weather during the peri-partum period, are probably important determinants of the observed seasonality.

In the current study, hazards of CM occurrence after 13 DIM were fairly similar for cows calving during the winter, spring, and summer seasons. Cows calving between September 21st and December 20th, however, showed lower hazard of CM occurrence after 13 DIM when

compared to cows calving during winter time (HR: 0.72; 95% CI: 0.60, 0.87). This lower risk of CM in the remaining lactation can potentially be explained by the same mechanisms previously described, since autumn calving cows will spend an important part of their remaining lactation beyond 13 DIM in cooler weather (i.e. winter) and will be in a relatively more advanced state of their lactation when the warmer summer temperature begins. In contrast, winter calving cows are assured to spend a substantial part of their first few months in milk in warm and often humid weather.

3.6.1.4. Herd somatic cell score

In the current study, hazard of CM in the first 13 DIM increased with decreasing herd SCS. At the quarter-level, evidence suggesting either a "protective" (Schukken et al., 1994; Schukken et al., 1999; Suriyasathaporn et al., 2000) or "causal" (Green et al., 2007; Pantoja et al., 2008; Steeneveld et al., 2008) effect of higher quarter SCC against IMI has been reported. At herd-level, however, lower mean herd SCS and higher proportion of cows with low SCC have both been associated with higher risk of CM (Erskine et al., 1988; Beaudeau et al., 2002; de Haas et al., 2005). In the data presented here, the contextual effect (i.e. the effect of the herd SCS on a specific cow) potentially operates through a different biological process than the individual effect (i.e. the effect of the cow's own previous SCS). Herds having lower SCS have usually achieved a certain level of control of contagious pathogens such as *Staphylococcus aureus* and *Streptococcus agalactiae,* and better control of these contagious pathogens would be expected to reduce the absolute number and proportion of clinical cases associated with these specific organisms. In low SCS herds, it has been shown that environmental pathogens are more frequently cultured from CM cases (Erskine et al., 1988; Barkema et al., 1998) and that CM cases are most often observed in these herds during the first month of lactation (Erskine et al., 1988). This later finding could be indicative of

infections acquired during the dry-period or early lactation infection, typical features of most environmental pathogens. Since CM cases caused by contagious pathogens tend to be more evenly spread over the lactation, a shift toward CM occurring mainly in the early lactating period in herds that have efficiently controlled contagious pathogens is not surprising. The increased hazard of CM during the early lactating period observed in the current and in previous studies (Erskine et al., 1988; de Haas et al., 2005), however, remains to be completely elucidated. It seems rather unlikely that a better control of contagious pathogens at the herd level would actually result in a higher absolute number of CM cases. Shuster et al. (1996) and Vandeputte-Van Messom et al. (1993), however, both demonstrated an increase in mastitis severity in quarters with lower SCC. Mastitis severity, in turn, will directly influence mastitis detection. Moreover, mastitis severity has been reported to strongly influence treatment decisions (Vaarst et al., 2002; Dufour et al., 2010), and administration of an antimicrobial treatment will certainly influence reporting of a CM case. There is, therefore, a strong possibility that the observed association between herd SCS and CM hazard is the result of an increased severity of CM cases due to a higher proportion of quarters with higher susceptibility (i.e. with lower SCC), and that this shift toward more severe cases resulted in increased detection and/or reporting of CM cases, rather than an absolute increase in the number of cases.

3.6.1.5. Parity, pasture access, and cases of CM treated with antimicrobials

In the current study, we observed a higher CM incidence between 0-13 DIM in first parity cows confined inside compared to second lactation cows housed similarly. In comparison, in herds where cows had access to pasture, we observed a relatively straightforward relationship between parity and CM incidence during the 0-13 DIM and >13 DIM periods, with lower CM hazard in first parity cows and CM incidence increasing with every additional lactation. This latter consistent relationship of increasing CM hazard with increasing parity has also

been reported before in studies conducted on confined dairy cows, but for which the outcome of interest was CM occurrence over the whole lactation (Gröhn et al., 2004; Hertl et al., 2011) or CM occurrence during the first 30 and 60 days of lactation (Green et al., 2007; Pinedo et al., 2012). Conversely, studies conducted on confined dairy cows, but focusing precisely on CM occurrence during the first 2 weeks in milk, reported higher CM incidence during this period for first lactation cows compared to older cows (Barkema et al., 1998; Steeneveld et al., 2008; Olde Riekerink et al., 2008). Observations from the current study (i.e. higher CM hazard in confined heifers compared to older cows in the 0-13 DIM period followed by increasing hazard by parity in the >13 DIM period) suggests a protective effect of pasture for 1st parity cows, and this is consistent with the literature. The same observation has been made by Waage et al. (1998) who reported lower CM risk when heifers had access to pasture around calving or during their post-partum period. In both the Waage et al. (1998) and the current studies, only heifers that actually had access to pasture during the early lactating period did benefit from this practice. As expected, first lactation cows calving during winter (i.e. confined around calving time) in herds where cows are sent to pasture during summer did not show lower 0-13 DIM CM hazard.

The generally increasing CM hazard with increasing parity is rather straightforward to explain. Older cows have been exposed to multiple pathogens over a long time period which could result in a higher proportion of subclinically infected quarters. Clinical flare-up of these infections could then yield more CM and often recurrence of CM. Furthermore, because of the long exposure to milking machines, older cows are more likely to have more callous and rough teat ends which can also result in higher risk of CM (Neijenhuis et al., 2000 and 2001).

The 0-13 DIM CM hazard difference observed between housed and pastured heifers is particularly interesting. The lower 0-13 DIM CM hazard in pastured heifers is likely to be the result of the decreased pressure of infection from their environment compared to the

environment of housed heifers. Moreover, in most Canadian dairies, heifers are usually moved with a new group of cows around calving where they have to adapt to a new facility and a new diet, compete with older and heavier cows for feed, water, and stalls, and establish a new social network. The stress resulting from these different adaptations can pay its toll on the first lactation cow's immune system and make them more prone to infectious diseases such as mastitis. Even when compared with cows housed in well-designed facilities, the lower animal density and the environment found at pasture is more likely to reflect the natural cow environment and to minimize the stress associated with the post-partum period.

Finally, in the current study, treating more than 50% of CM cases had very little effect on CM occurrence in the first 13 DIM, but was associated with slightly increased CM hazard after 13 DIM in older cows. This observation could actually result from an inappropriate balance between treating and culling mastitic cows. In the current study, 71.0% of the producers classified as treating ≥ 50% of cases, actually reported treating almost all cases (i.e. ≥ 90%; see Dufour et al., 2010). We can hypothesize that culling rather than treating a certain proportion of these cases would have reduced the risk of transmission of chronic and well host-adapted pathogens to other herdmates; hence the higher CM hazard observed in herds where a large proportion of the cases are treated. Furthermore, we cannot exclude a potential association between proportions of CM cases treated and reported. Producers that are convinced that all CM cases need treatment may be more aggressive in trying to detect and, perhaps, report these cases.

3.6.1.6. Strengths and limitations of current study

Risk factors for clinical mastitis have been widely investigated in a number of locations and under a variety of management systems. Overall, most of our results are broadly consistent with previous results. However, this study had a number of features that make the results

important in terms of our understanding of mastitis. First, the size of the data set (over 10,000 lactations from 69 herds) makes it one of the largest datasets assembled for mastitis research which will have contributed to more precise estimates of effects. Second, the analytical techniques used allowed us to clearly separate the nature of the effects in the early post-partum period (0-13 days) and later, while still appropriately accounting for the clustered nature of the data. A limitation of the study was that most of the factors examined were herd-level factors and the study contained only 69 herds. These herds were chosen to be representative of the Canadian dairy population (Reyher et al., 2011) but, given the intensive nature of the data collection required, random sampling of the population was not possible.

3.6.2. Analysis of simulation study

The simulation study compared the performance of PML approach with either adaptive Gaussian quadrature or Laplace approximation with the penalized partial likelihood approach (PPL) in terms of the bias of the point estimates, their empirical variability, and the bias of the estimation of such variability. The results showed that the Poisson likelihood approach with adaptive Gaussian quadrature performed well in all regards and produced nearly unbiased estimates for model parameters, including cluster variance estimates and their standard errors. However, the Poisson approach with Laplace approximation tended to strongly overestimate the between-cow variance and underestimate the between-herd variance, as well as produced estimates with downward bias for the standard errors. This later finding implies a higher than expected type I error rate when using Laplace approximation in a similarly structured dataset. The overestimation of between-subcluster variance was also reported in Feng et al. (2009). As Feng et al. (2009) used a subcluster size of 2 and in our case it was even smaller (mean 1.348, range 1-3), such bias is probably attributable to the small subcluster size and the asymptotic nature of the Laplace approximation that requires

reasonably large cluster size (Joe, 2008). The performance of PPL was good and comparable to the PMLGQ approach in estimating the fixed effect parameters and their standard errors, but the approach tended to underestimate the between-cow variance, this is again probably due to the use of Laplace approximation for a small subcluster size. This bias in variance component estimates of PPL was also pointed out in Pankratz et al. (2005).

3.6.3. Choice of estimation approach

Estimation techniques for hierarchical Cox models are not straightforward. The current implementation of nested frailty models, such as those implemented in coxme (Therneau, 2013), frailtyHL (Ha et al., 2012), and frailtypack (Rondeau et al., 2012) libraries of R software, are still limited to models with few predictors and moderate size of datasets. Both the coxme and the frailtyHL implementations of nested log-normal frailty models do not support the counting process data format necessary for time-dependent predictors and effects. These approaches are also based only on Laplace approximations which appear to perform suboptimally in nested frailty models with small subcluster sizes. On the other hand, the frailtyPenal function of the frailtypack package assumes a gamma distribution for nested frailties and can deal with counting process formatted data, but it requires the number of random effects to be at most moderate. This effectively precluded its use for our data with 8,035 cows from 69 herds. This study demonstrates that the above discussed limitations of nested frailty model implementations can be overcome by reformulating the nested frailty Cox model as a nested random-effects Poisson model and using standard GLMM software for estimation.

3.7. Conclusions

In summary, analyzing large survival datasets with multiple levels of clustering requires accounting for the correlation between event times within each of these levels, as well as

handling the time-dependent variables and effects that often present in the data. A Poisson modeling approach with adaptive Gaussian quadrature provided fairly robust estimation for Cox models with nested log-normal frailty while the penalized partial likelihood and the Poisson maximum likelihood with Laplacian approximation were found to have substantial drawbacks. The research indicated that some of the herd managerial factors combined with cow characteristics influence the hazard of CM during the lactation period; some of these effects were different very early than later in the lactation.

Acknowledgements

The authors acknowledge the Canadian Bovine Mastitis Research Network for providing the CBMRN data and the Atlantic Computational Excellence Network (ACEnet) for access to computational facilities.

References

Barkema H, Schukken Y, Lam T, Beiboer M, Wilmink H, Benedictus G, Brand A. (1998). Incidence of clinical mastitis in dairy herds grouped in three categories by bulk milk somatic cell counts. *Journal of Dairy Science*, **81**: 411-419.

Barkema H, Schukken Y, Lam T, Beiboer M, Benedictus G, Brand A. (1999). Management practices associated with the incidence rate of clinical mastitis. *Journal of Dairy Science*, **82**: 1643-1654.

Barnouin J, Bord S, Bazin S, Chassagne M. (2005). Dairy management practices associated with incidence rate of clinical mastitis in low somatic cell score herds in France. *Journal of Dairy Science*, **88**: 3700-3709.

Beaudeau F, Fourichon C, Seegers H, Bareille N. (2002). Risk of clinical mastitis in dairy herds with a high proportion of low individual milk somatic-cell counts. *Preventive Veterinary Medicine*, **53**: 43-54.

Bender R, Augustin T, Blettner M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, **24**: 1712-1723.

Borne B, Vernooij J, Lupindu A, Schaik G, Frankena K, Lam T, Nielen M. (2011). Relationship between somatic cell count status and subsequent clinical mastitis in Dutch dairy cows. *Preventive Veterinary Medicine*, **102**: 265-273.

Cox D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*, **34**: 187-220.

De Haas Y, Barkema H, Schukken Y, Veerkamp R. (2005). Associations between somatic cell count patterns and the incidence of clinical mastitis. *Preventive Veterinary Medicine*, **67**: 55-68.

Dohoo I, Martin W, Stryhn H. (2009). *Veterinary Epidemiologic Research*. 2nd edition, Charlottetown, VER - Inc.

Duchateau L, Janssen P. (2008). *The Frailty Model*. Springer-Verlag.

Ducrocq V, Casella G. (1996). A Bayesian analysis of mixed survival models. *Genetics Selection Evolution*, **28**: 505-529.

Dufour S, Barkema H, DecCoteaux L, DeVries T, Dohoo I, Reyher K, Roy J, Scholl D. (2010). Development and validation of a bilingual questionnaire for measuring udder health related management practices on dairy farms. *Preventive Veterinary Medicine*, **95**: 74-85.

Elbers A, Milteburg J, De Lange D, Crauwels A, Barkema H, Schukken Y. (1998). Risk factors for clinical mastitis in a random sample of dairy herds from the southern part of The Netherlands. *Journal of Dairy Science*, **81**: 420-426.

Erskine R, Eberhart R, Hutchinson L, Spencer S, Campbell M. (1988). Incidence and types of clinical mastitis in dairy herds with high and low somatic cell counts. *Journal of the American Veterinary Medical Association*, **192**: 761-765.

Feng S, Wolfe R, Port F. (2005). Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *Journal of the American Statistical Association*, **100**: 718-735.

Feng S, Nie L, Wolfe R. (2009). Laplace's approximation for relative risk frailty models. *Lifetime Data Analysis*, **15**: 343-356.

Gasqui P, Barnouin J. (2003). Statistical modelling for clinical mastitis in dairy cow: problems and solutions. *Veterinary Research*, **34**: 493-505.

Gasqui P, Coulon J, Pons O. (2003). An individual modelling tool for within and between lactation consecutive cases of clinical mastitis in the dairy cow: an approach based on a survival model. *Veterinary Research*, **34**: 85-104.

Goethals K, Ampe B, Berkvens D, Laevens H, Janssen P, Duchateau L. (2009). Modelling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**: 1-14.

Green M, Green L, Medley G, Schukken Y, Bradley A. (2002). Influence of dry period bacterial intramammary infection on clinical mastitis in dairy cows. *Journal of Dairy Science*, **85**: 2589-2599.

Green M, Bradley A, Medley G, Browne W. (2007). Cow, farm, and management factors during the dry period that determine the rate of clinical mastitis after calving. *Journal of Dairy Science*, **90**: 3764-3776.

Gröhn Y, Wilson D, Gonzalez R, Hertl J, Schulte H, Bennett G, Schukken Y. (2004). Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. *Journal of Dairy Science*, **87**: 3358-3374.

Gröhn Y, Gonzalez R, Wilson D, Hertl J, Bennett G, Schulte H, Schukken Y. (2005). Effect of pathogen-specific clinical mastitis on herd life in two New York State dairy herds. *Preventive Veterinary Medicine*, **71**: 105-125.

Ha I, Noh M, Lee Y. (2012). FrailtyHL: A package for fitting frailty models with H-likelihood. *The R Journal*, **4**: 28-37.

Hangagal D. (2011). *Modelling Survival Data Using Frailty Models*. Chapman & Hall/CRC.

Hertl J, Schukken Y, Bar D, Bennett G, Gonzalez R, Rauch B, Welcome F, Tauer L, Gröhn Y. (2011). The effect or recurrent episodes of clinical mastitis caused by gram-positive and gram-negative bacteria and other organisms on mortality and culling in Holstein dairy cows. *Journal of Dairy Science*, **94**: 4863-4877.

Joe H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*, **52**: 5066-5074.

Ma R, Krewski D, Burnett R. (2003). Random effects Cox models: A Poisson modelling approach. *Biometrika*, **90**: 157-169.

Neijenhuis F, Barkema H, Hogeveen H, Noordhuizen J. (2000). Classification and longitudinal examination of callused teat ends in dairy cows. *Journal of Dairy Science*, **83**: 2795-2804.

Neijenhuis F, Barkema H, Hogeveen H, Noordhuizen J. (2001). Relationship between teat-end callosity and occurrence of clinical mastitis. *Journal of Dairy Science*, **84**: 2664-2672.

Nyman A, Ekman T, Emanuelson U, Gustafsson A, Holteniusk K, Waller K, Sandgren C. (2007). Risk factors associated with the incidence of veterinary-treated clinical mastitis in Swedish dairy herds with a high milk yield and a low prevalence of subclinical mastitis. *Preventive Veterinary Medicine*, **78**: 142-160.

Olde Riekerink R, Barkema H, Kelton D, Scholl D. (2008). Incidence rate of clinical mastitis on Canadian dairy farms. *Journal of Dairy Science*, **91**: 1366-1377.

O'Reilly K, Green M, Peeler E, Fitzpatrick J, Green L. (2006). Investigation of risk factors for clinical mastitis in British dairy herds with bulk milk somatic cell counts less than 150,000 cells/ml. *Veterinary Record*, **158**: 649-653.

Pankratz S, Andrade M, Therneau T. (2005). Random-effects Cox proportional hazards model: General variance components methods for time-to-event data. *Genetic Epidemiology*, **28**: 97-109.

Pantoja J, Hulland C, Ruegg P. (2009). Somatic cell count status across the dry period as a risk factor for the development of clinical mastitis in the subsequent lactation. *Journal of Dairy Science*, **92**: 139-148.

Pinedo P, Fleming C, Risco C. (2012). Events occurring during the previous lactation, the dry period, and peripartum as risk factors for early lactation mastitis in cows receiving 2 different intramammary dry cow therapies. *Journal of Dairy Science*, **95**: 7015-7026.

Rabe-Hesketh S, Skrondal A. (2012). *Multilevel and Longitudinal Modeling using Stata*. *Volume II: Categorical Responses, Counts, and Survival*, 3rd ed, Stata Press.

Reyher K, Dufour S, Barkema H, DecCoteaux L, DeVries T, Dohoo I, Keefe G, Roy J, Scholl D. (2011). The national cohort of dairy farms—A data collection platform for mastitis research in Canada. *Journal of Dairy Science*, **94**: 1616-1626.

Richert R, Cicconi K, Gamroth M, Schukken Y, Stiglbauer K, Ruegg P. (2013). Risk factors for clinical mastitis, ketosis, and pneumonia in dairy cattle on organic and small conventional farms in the United States. *Journal of Dairy Science*, **96**: 4269-4285.

Ripatti S, Palmgren J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56**: 1016-1022.

Rondeau V, Filleul L, Joly P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine*, **25**: 4036-4052.

Rondeau V, Mazroui Y, Gonzalez J. (2012). frailtypack: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, Volume 47, Issue 4.

Rupp R, Boichard D. (2000). Relationship of early first lactation somatic cell count with risk of subsequent first clinical mastitis. *Livestock Production Science*, **62**: 169-180.

Santos J, Cerri R, Ballou M, Higginbotham G, Kirk J. (2004). Effect of timing of first clinical mastitis occurrence on lactational and reproductive performance of Holstein dairy cows. *Animal Reproduction Science*, **80**: 31-45.

Sastry N. (1997). A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of the American Statistical Association*, **92**: 426-435.

Schepers J, Dijkhuizen A. (1991). The economics of mastitis and mastitis control in dairy cattle: a critical analysis of estimates published since 1970. *Preventive Veterinary Medicine*, **10**: 213-224.

Schukken Y, Mallard B, Dekkers J, Leslie K, Stear M. (1994). Genetic impact on the risk of intramammary infection following *Staphylococcus aureus* challenge. *Journal of Dairy Science*, **77**: 639-647.

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

Schukken Y, Bar D, Hertl J, Gröhn Y. (2010). Correlated time to event data: Modelling repeated clinical mastitis data from dairy cattle in New York State. *Preventive Veterinary Medicine*, **97**: 150-156.

Shuster D, Lee E, Kehrli M. (1996). Bacterial growth, inflammatory cytokine production, and neutrophil recruitment during coliform mastitis in cows within ten years after calving, compared with cows at midlactation. *American Journal of Veterinary Research*, **57**: 1569-1575.

Steeneveld W, Hogeveen H, Barkema H, Van den Broek J, Huirne R. (2008). The influence of cow factors on the incidence of clinical mastitis in dairy cows. *Journal of Dairy Science*, **91**: 1391-1402.

Suriyasathaporn W, Schukken Y, Nielen M, Brand A. (2000). Low somatic cell count: a risk factor for subsequent clinical mastitis in a dairy herd. *Journal of Dairy Science*, **83**: 1248-1255.

Therneau T, Grambsch P. (2000). *Modelling Survival Data, Extending the Cox model*. New York: Springer.

Therneau T. (2013). 'coxme'. A package for mixed effects Cox models. R package version 2.2-3, URL http://cran.r-project.org/web/packages/cox me/.

Therneau T, Crowson C. (2013). Using time-dependent covariates and time-dependent coefficients in the Cox model. The Survival Package (R help guide).

Vaarst M, Paarup-Laursen B, Houe H, Fossing C, Andersen H. (2002). Farmers' choice of medical treatment of mastitis in Danish dairy herds based on qualitative research interviews. *Journal of Dairy Science*, **85**: 992-1001.

Vandeputte-Van Messom G, Burvenich C, Roets E, Massart-Leen A, Heyneman R, Kremer W, Brand A. (1993). Classification of newly calved cows into moderate and severe responders to experimentally induced Escherichia coli mastitis. *Journal of Dairy Research*, **60**: 19-29.

Waage S, Sviland S, Odegaard S. (1998). Identification of risk factors for clinical mastitis in dairy heifers. *Journal of Dairy Science*, **81**: 1275-1284.

Wienke A. (2010). *Frailty Models in Survival Analysis*. Chapman & Hall/CRC.

Table 3.1. Explanatory variables used in the analyses of the January 2007-December 2008 clinical mastitis data from the Canadian Bovine Mastitis Research Network (CBMRN).

Variable	Description
Geographic region	With 4 categories: Alberta, Ontario, Quebec, Atlantic; herd level.
Calving season	With 4 categories: winter (21 Dec.-20 Mar.), spring (21 Mar.-20 Jun.), summer (21 Jun.-20 Sep.), autumn (21 Sep.-20 Dec.); lactation level.
Cow parity	With 3 categories; 1st lactation, 2nd lactation, and ≥3rd lactation cows; lactation level.
Mean herd somatic cell score	Continuous: mean herd somatic cell score (SCS) during the previous sampling period; herd level.
Number of milking cows in a herd*	Continuous: mean of 6-month period prior the current period; herd level.
Housing type	Trichotomous: tie stall, free stall and bedding pack, herd level.
Milking procedures	
Wear gloves during milking*	Dichotomous: yes/no; herd level.
Fore-stripping*	Dichotomous: yes/no; herd level.
Pre-milking teat disinfection	Dichotomous: yes/no; herd level.
Post milking teat disinfection	Dichotomous: yes/no; herd level.
Environment	
Material used for base of stalls or pens*	Trichotomous: concrete, sand, and mattress or rubber mat; herd level.
Type of bedding used in stalls or pens	Trichotomous: wood, sand, and straw; herd level.
Cows have access to pasture*	Dichotomous: yes/no; herd level.
Dry-off period management	
Use external teat sealant at dry-off (ETS)	Dichotomous: yes/no; herd level.
Use internal teat sealant at dry-off (ITS)	Dichotomous: yes/no; herd level.
Proportion of cows receiving antimicrobial treatment at drying off (DCT)	Dichotomous: 100% vs. <100%; herd level.
Vaccination against coliforms	Dichotomous: yes/no; herd level.
Proportion of calving occurring in maternity pen*	Dichotomous: low (≤ 50%) vs. high (> 50%); herd level.
Others	
Proportion of clinical mastitis (CM) cases treated*	Dichotomous: low (≤ 50%) vs. high (> 50%); herd level.
Udder hair management	Dichotomous: yes (clipped or flamed)/no; herd level.
Tail management*	Dichotomous: yes (clipped or tied)/no; herd level.

* Time-varying predictor.

Table 3.2. Descriptive statistics for lactation-level predictors in the CBMRN data from 69 herds between January 2007 and December 2008.

Predictor		Number of lactations	Proportion of lactations with a clinical mastitis event
Parity	1st lactation	3629	0.034
	2nd lactation	2970	0.041
	≥3rd lactation	4232	0.067
Calving season	winter	2706	0.042
	spring	2521	0.038
	summer	2831	0.039
	autumn	2773	0.023

Table 3.3. Descriptive statistics for herd-level predictors in the CBMRN data from 69 herds between January 2007 and December 2008.

Predictor		Number of herd 6-month periods	Proportion of lactations with a clinical mastitis event
Region	Alberta	37	0.019
	Ontario	96	0.055
	Quebec	78	0.033
	Atlantic	54	0.035
Housing type	Tie-stall	180	0.088
	Free-stall	77	0.050
	Bedding-pack	8	0.004
Milking procedures			
Wear gloves during milking*	No	111	0.057
	Yes	150	0.085
Fore-stripping*	No	109	0.055
	Yes	156	0.086
Pre-milking teat disinfection	No	84	0.037
	Yes	176	0.104
Post milking teat disinfection	No	4	0.003
	Yes	256	0.139
Environment			
Material used for base of stalls or pens*	Concrete	31	0.010
	Sand	25	0.016
	Mattress or rubber mat	209	0.116
Type of bedding used in stalls or pens	Wood	54	0.030
	Sand	10	0.011
	Straw	201	0.101
Pasture access*	No	201	0.106
	Yes	64	0.036
Dry-off period management			
Use external teat sealant at dry-off (ETS)	No	237	0.129
	Yes	23	0.013
Use internal teat sealant at dry-off (ITS)	No	157	0.129
	Yes	106	0.013
Prop. of cows receiving antimicrob. treat. at dry-off	< 100%	40	0.022
	= 100%	220	0.120
Vaccination	No	147	0.072
	Yes	113	0.069
% of calving occurring in maternity pen*	≤ 50%	115	0.061
	> 50%	150	0.081
Other management			
Proportion of clinical mastitis (CM) cases treated*	≤ 50%	77	0.032
	> 50%	188	0.110
Udder hair management	No	49	0.024
	Flamed or clipped	216	0.118
Tail management*	No	55	0.030
	Clipped or tied	210	0.112

Predictor	Mean (sd) (cases)	Mean (sd) (non-cases)
Mean of herd somatic cell score (SCS) in previous period	2.66 (0.494)	2.61 (0.507)
Number of milking cows in a herd (mean of 6-month period)*	89.0 (44.8)	84.4 (45.7)

*Time-varying predictor.

Table 3.4. Parameter estimates, standard errors and P-values in the final nested frailty Cox model of CBMRN data between January 2007 and December 2008.

Predictor/Parameter[a]		Estimate (SE)	Overall[b] P-value	Time[c] component (SE)	Overall[d] P-value
Fore-stripping	Yes vs. no	0.336 (0.178)	0.059		
Calving season	Spring vs. Winter	-0.043 (0.081)	0.004	0.687 (0.198)	0.000
	Summer vs. Winter	-0.123 (0.084)		1.030 (0.194)	
	Autumn vs. Winter	-0.327 (0.093)		0.434 (0.221)	
Mean of herd SCS		0.058 (0.123)	0.638	-0.465 (0.152)	0.002
Pasture access	Yes vs. no	-0.188 (0.170)	0.269	-0.828 (0.345)	0.017
% of cases of CM treated	> 50% vs. ≤ 50%	-0.008 (0.221)	0.969	-0.336 (0.150)	0.025
Parity	2nd lactation vs. 1st lactation	0.343 (0.167)	0.015	-1.167 (0.210)	0.000
	≥3rd lactation vs. 1st lactation	0.443 (0.156)		-0.696 (0.172)	
Interactions					
	2nd lactation × pasture access (yes)	0.043 (0.197)	0.044	1.078 (0.465)	0.032
	≥3rd lactation × pasture access (yes)	0.375 (0.182)		0.956 (0.394)	
	2nd lactation × % of treated cases of CM (> 50%)	0.373 (0.176)	0.052		
	≥3rd lactation × % of treated cases of CM (> 50%)	0.350 (0.161)			
Region	Ontario vs. Western Canada	0.358 (0.259)	0.277		
	Quebec vs. Western Canada	-0.018 (0.293)			
	Atlantics vs. Western Canada	0.199 (0.321)			
Between-cow variance		0.498 (0.140)			
Between-herd variance		0.394 (0.086)			

[a] Coefficients for the 4th order polynomial function of time represents the baseline hazard not shown.

[b] Overall P-value for main effect or interaction with other predictor (after 13 DIM, if involved in time component).

[c] Time modelled as two time periods within lactation (1: first 13 DIM vs. 0: after 13 DIM); estimates shown are interaction terms between time component and effect.

[d] Overall P-value for time components (i.e. interaction term).

Table 3.5: Estimated hazard ratios, their standard errors and 95% confidence intervals for calving season and the combinations of pasture access, proportion of cases of CM treated, and cow parity; in the first 13 DIM and after 13 DIM of CBMRN data between Jan. 2007 and Dec. 2008.

	First 13 DIM		After 13 DIM	
Predictor	**Hazard ratio (SE)**	**95% CI**	**Hazard ratio (SE)**	**95% CI**
Calving season				
Spring vs. winter	1.906 (0.353)	(1.326, 2.740)	0.958 (0.077)	(0.817, 1.122)
Summer vs. winter	2.477 (0.448)	(1.738, 3.531)	0.884 (0.074)	(0.751, 1.104)
Autumn vs. winter	1.113 (0.228)	(0.744, 1.664)	0.721 (0.067)	(0.600, 0.866)
Summer vs. spring	1.301 (0.193)	(0.973, 1.739)	0.923 (0.079)	(0.780, 1.092)
Autumn vs. spring	0.584 (0.104)	(0.413, 0.828)	0.753 (0.073)	(0.623, 0.910)
Autumn vs. summer	0.449 (0.073)	(0.327, 0.618)	0.815 (0.080)	(0.673, 0.988)
Pasture access				
1^{st} lactation	0.361 (0.117)	(0.192, 0.683)	0.829 (0.141)	(0.594, 1.156)
2^{nd} lactation	1.112 (0.339)	(0.611, 2.022)	0.865 (0.131)	(0.644, 1.165)
3^{rd} + lactation	1.371 (0.270)	(0.933, 2.016)	1.206 (0.153)	(0.940, 1.547)
Proportion of CM treated				
1^{st} lactation	0.708 (0.172)	(0.441, 1.139)	1.009 (0.223)	(0.643, 1.531)
2^{nd} lactation	1.029 (0.258)	(0.630, 1.683)	1.440 (0.313)	(0.941, 2.204)
3^{rd} + lactation	1.006 (0.233)	(0.639, 1.584)	1.408 (0.288)	(0.943, 2.102)
Cow parity				
2^{nd} lactation vs. 1^{st} lactation				
≤ 50% treated CM & no past. acces.	0.439 (0.102)	(0.261, 0.739)	1.409 (0.235)	(1.016, 1.955)
≤ 50% treated CM & past. acces.	1.347 (0.536)	(0.618, 2.938)	1.471 (0.318)	(0.964, 2.246)
> 50% treated CM & no past. acces.	0.637 (0.124)	(0.435, 0.934)	2.047 (0.207)	(1.678, 2.495)
> 50% treated CM & past. acces.	1.957 (0.752)	(0.921, 4.155)	2.138 (0.386)	(1.500, 3.042)
3^{rd} + lactation vs. 1^{st} lactation				
≤ 50% treated CM & no past. acces.	0.777 (0.150)	(0.532, 1.133)	1.558 (0.243)	(1.148, 2.113)
≤ 50% treated CM & past. acces.	2.941 (0.996)	(1.515, 5.713)	2.267 (0.441)	(1.547, 3.319)
> 50% treated CM & no past. acces.	1.102 (0.171)	(0.814, 1.494)	2.212 (0.214)	(1.831, 2.673)
> 50% treated CM & past. acces.	4.175 (1.375)	(2.189, 7.960)	3.218 (0.538)	(2.319, 4.468)
3^{rd} + lactation vs. 2^{nd} lactation				
≤ 50% treated CM & no past. acces.	1.770 (0.395)	(1.142, 2.743)	1.106 (0.169)	(0.982, 1.244)
≤ 50% treated CM & past. acces.	2.183 (0.662)	(1.205, 3.956)	1.541 (0.281)	(1.273, 1.863)
> 50% treated CM & no past. acces.	1.730 (0.326)	(1.196, 2.502)	1.081 (0.093)	(0.913, 1.278)
> 50% treated CM & past. acces.	2.134 (0.613)	(1.215, 3.748)	1.506 (0.223)	(1.126, 2.013)

Table 3.6. Results of simulation study based on a log-normal nested frailty Cox model of CBMRN data between January 2007 and December 2008. Mean of the estimate, empirical standard deviation, mean of the model-based standard error, bias and probability coverage over 250 simulated datasets.

		Lactation-level		Cow-level	Herd-level			Convergence
Method		β_1	β_2	β_3	β_4	σ_c^2	σ_h^2	rate
	True value	0.4	0.6	0.5	-0.8	0.5	0.3	
PPL	**Estimate**	0.386	0.578	0.477	-0.758	0.342	0.287	100%
	Emp. Std	0.078	0.071	0.053	0.138	0.085	0.065	
	Model Se	0.074	0.068	0.057	0.146	--[a]	--[a]	
	Relative bias	-0.035	-0.037	-0.046	0.053	-0.316	-0.043	
	95% CI cover.	92%	94%	94%	96%	--[a]	--[a]	
PMLAP	**Estimate**	0.418	0.630	0.517	-0.835	1.492	0.199	72%
	Emp. Std	0.094	0.090	0.070	0.173	1.699	0.126	
	Model Se	0.079	0.075	0.064	0.128	0.214	0.047	
	Relative bias	0.045	0.050	0.034	0.044	1.984	-0.367	
	95% CI cover.	91%	90%	94%	84%	75%	100%	
PMLGQ	**Estimate**	0.399	0.598	0.491	-0.791	0.482	0.305	98%
	Emp. Std	0.080	0.074	0.056	0.144	0.155	0.089	
	Model Se	0.076	0.071	0.059	0.148	0.134	0.065	
	Relative bias	-0.003	-0.003	-0.018	0.011	-0.036	0.017	
	95% CI cover.	93%	95%	95%	96%	96%	96%	

PPL: Penalized partial likelihood; PMLAP: Poisson maximum likelihood with Laplace approximation; PMAGQ: Poisson maximum likelihood with adaptive Gaussian quadrature.

[a] No available estimate (current implementation of PPL procedure in R software does not provide standard errors for variance estimates).

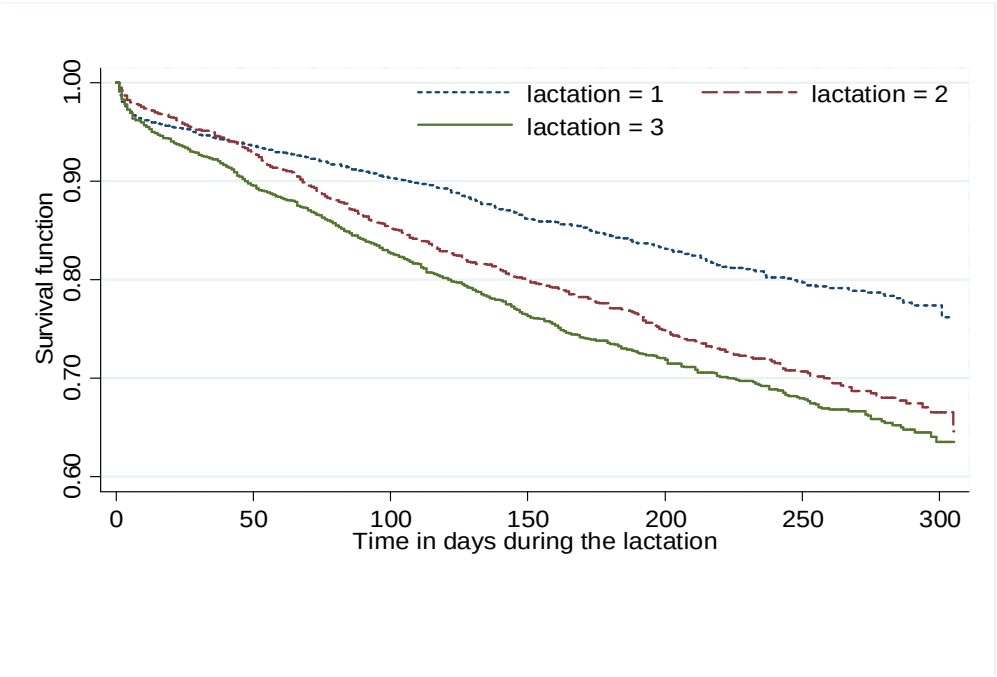

Figure 3.1: Kaplan-Meier survivor curves for clinical mastitis events up till 305 DIM for 1st, 2nd, and ≥3rd lactation cows of CBMRN data between January 2007 and December 2008. The short-dashed, long-dashed, and solid patterns are 1st, 2nd, ≥3rd lactation cows, respectively.

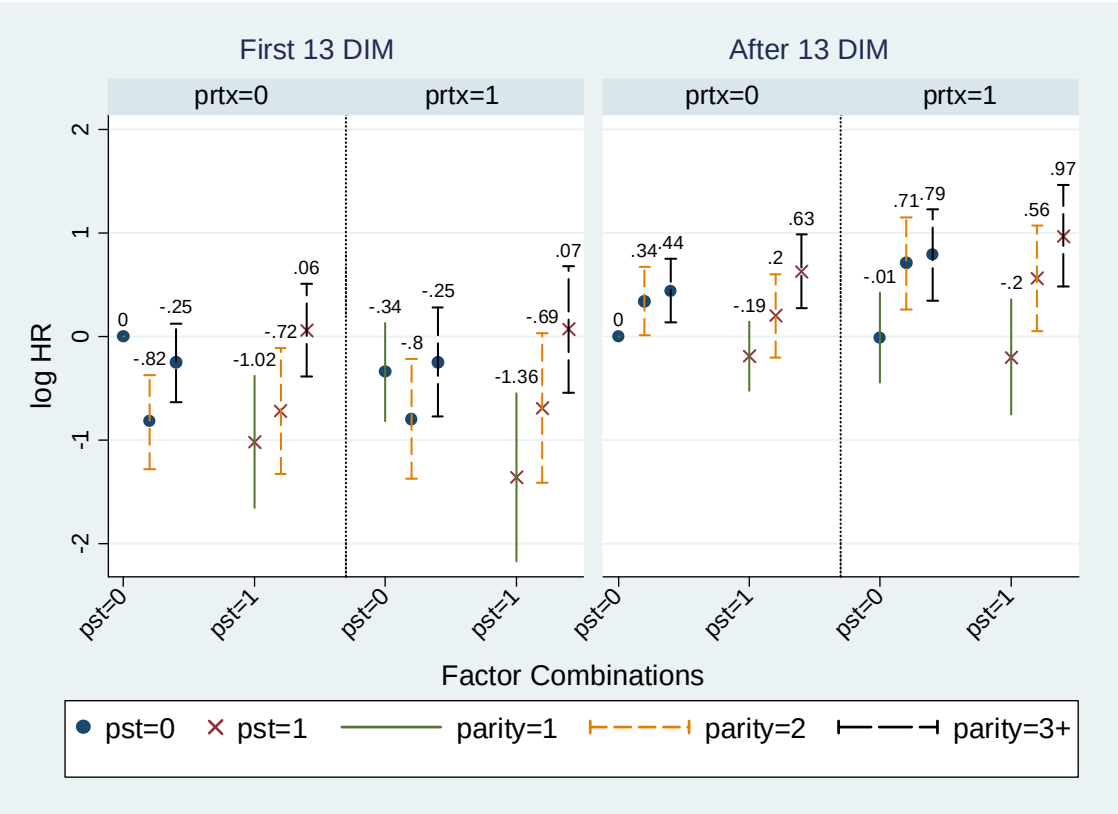

Figure 3.2: The log of hazard ratios of clinical mastitis and their confidence intervals for the combinations of cow parity, pasture access (pst), and proportion of cases of CM treated (prtx), in the first and after 13 DIM of CBMRN data between January 2007 and December 2008. The first four triples are for effects within the first 13 DIM and the second four triples are for effects after 13 DIM. Within a group of four triples, the first two triples represent proportion of cases of CM treated ≤ 50% and next two triples represent proportion of cases of CM treated > 50%. Within a group of two triples, the first triple is for no pasture access and the second triple is for pasture access. Solid lines are for 1st lactation cows, dashed lines are for 2nd lactation cows, and long-dashed lines are for ≥3rd lactation cows.

Chapter 4

A cross-classified and multiple membership Cox model applied to calf mortality data

The work of this chapter is published as: Elghafghuf A, Stryhn H, Waldner C. (2014). A cross-classified and multiple membership Cox model applied to calf mortality data. *Preventive Veterinary Medicine*, 115: 29-38.

4.1. Abstract

A cross-classified and multiple membership Cox model was applied to calf mortality data from Western Canada, where 23,409 calves from 174 herds were followed for up to 180 days after calving. The herds were cross-classified by 49 veterinary clinics and 9 ecological regions and in a multiple membership relation to the veterinary clinics, resulting in a 3-level cross-classified and multiple membership data structure. The model was formulated in a mixed-effects Poisson model framework with normally distributed random effects, and was fitted to the data by Bayesian Markov Chain Monte Carlo (MCMC) estimation. Important fixed effects included whether the calf was a twin, calf gender, assistance at calving, cow age, average temperature the first week after calving, the percentage of the herd that had already calved, whether calf shelters were provided, whether cow-calf pairs were moved to a nursery area, and whether any animals were purchased into the herd at or near the time of calving. The analysis demonstrated a greater variation among herds than among both ecological regions and veterinary clinics. Further, a simulation study for a setting similar to the real data gave evidence that the used approach provides valid estimates.

4.2. Introduction

Researchers in veterinary epidemiology are often interested in modeling hierarchical data with a time-to-event response variable. Hierarchical time-to-event models, also referred to as hierarchical survival or frailty models, can be used with the nested data structures (e.g. animals nested within herds and herds located in different ecological regions) commonly found in veterinary science. One potential limitation of nested frailty models typically used for hierarchical survival data, however, is that they are designed to be used with perfectly hierarchical survival data, but in reality not all data structures found in the veterinary sciences are perfectly hierarchical. If in the previous example (some) herds are serviced by multiple

veterinary clinics, an imperfect hierarchical data structure is present where the lower level units (herds) are members of multiple higher level units (clinics) simultaneously. This structure is called a multiple membership data structure (Browne et al., 2001). In addition, different classifications may not be hierarchically nested in each other; in our example, herds serviced by a given clinic could be located in different ecological regions. This would mean that clinics are not hierarchically nested within regions, and the two factors should instead be viewed as (partially) cross-classified. In summary, the structure described corresponds to a 3-level cross-classified and multiple membership data structure (Browne et al., 2001).

Cross-classified and multiple membership (CMM) models have been proposed to account for such data structures (Browne et al., 2001; Fielding and Goldstein, 2006). The CMM model uses weights for multiple membership and takes into account cross-classified factors that might arise in the data. A few studies in veterinary epidemiology have used the CMM model with different response variables. Browne et al. (2001) applied a CMM model with a binary response to Danish poultry *Salmonella* outbreak data. Masaoud et al. (2010) fit a CMM logistic regression model to a dataset from aquaculture. Goldstein et al. (2002) introduced a linear response example of a multiple membership model for the milk yield of cows.

Many studies have shown that ignoring multiple membership or cross-classified data structure in the analysis can lead to invalid inference about the importance of the relevant data structure on the outcome of interest. For instance, Meyers and Beretvas (2006) and Luo and Kwok (2009) showed that ignoring one of the cross-classified factors in linear models results in biased estimation in the variance components and in the standard error of the regression coefficients. Results from Goldstein et al. (2007) demonstrated that using traditional models that ignore the multiple membership in the analysis when it is present underestimates the variance at the multiple membership level. In addition, models that take

into account the multiple membership structure give a better fit than models that ignore such structures. Recently, a simulation study conducted by Chung and Beretvas (2012) showed that ignoring multiple membership structure causes bias in the estimates of the regression coefficients and the variance component at the multiple membership level.

Despite the availability of veterinary data with CMM structure, few researchers have applied CMM models in veterinary medicine, and to our knowledge no studies have used a CMM model when the response variable is time-to-event. This could be due to complex estimation techniques for survival models with random effects (frailty models).

A review of the literature also suggested the need to re-examine the individual cow, herd management, and environmental factors associated with mortality in beef calves using a dataset with both detailed individual animal data and a relatively large number of herds. Many of the existing reports focus on calf loss at birth or in the perinatal period. While there are a number of observational studies published documenting calf loss after the perinatal period, most studies are either longitudinal studies from single research facilities focusing on individual animal attributes (Azzam et al., 1993; Patterson et al., 1987; Wittum et al., 1993), surveys with some data on individual animal attributes but a relatively small number of privately owned herds (Wittum et al., 1994; Ganaba et al., 1995), or herd level surveys with limited or no individual animal data (Schumann et al., 1990; Mathison, 1993; Dutil et al., 1999). Because previous studies have not taken the time to calf loss into account in the analysis, we found no reports to date that look objectively at when individual risk factors are of greatest risk to calf survival.

The first objective of the study is to explore and demonstrate the use of Poisson generalized linear mixed models (GLMMs) in Bayesian framework for estimating a Cox model with cross-classified and multiple membership frailties, and apply the approach to a large

observational dataset on calf mortality from veterinary medicine science. The second objective is to simultaneously examine the individual, herd management, and environmental factors associated with beef calf mortality in Western Canada and, where appropriate, to estimate the age period where calves are most at risk.

4.3. Materials and methods

4.3.1. Data

The data originated from the Western Canada beef productivity study (Waldner, 2008) which collected information on calf loss and mortality in beef cattle in Western Canada. We studied mortality in beef calves from January to June 2002 (180 days) which included a total of 24,647 calves and 971 cases of calf mortality from herds with complete local meteorological data. Calves with invalid values or missing information were excluded from the analysis. This eliminated less than 5% of observations including 74 cases of mortality. This strategy resulted in 23,409 calves, with 897 of these calves experiencing the event of interest. The event was defined as a case of calf mortality that happened at least one hour after birth; the event time was defined as the time from calving to death (recorded in days), and for those calves that died in the same day of birth the event time was set at 0.5. Calves that were sold during the follow-up period or survived until the end of the follow-up period were considered right censored observations. Because the observation period ended at the same time for all calves (June 30th), but all calves were born at different times during the calving season, we recognized the need for an analysis technique that accounted for different follow up times across the study population.

The dataset had a special hierarchical structure. In addition to calves being hierarchically nested within 174 herds, herds were cross-classified by 49 veterinary clinics and 9 ecological

regions (Waldner, 2008), and about 8% of the herds were registered in two veterinary clinics, resulting in a 3-level cross-classified and multiple membership data structure (Figure 4.1).

4.3.2. Statistical modeling

4.3.2.1. Frailty models for hierarchical survival data

Consider the example of 3-level hierarchical survival data with N animals from multiple herds and these herds located in different ecological regions. Let T_i and C_i denote the survival and censoring times, respectively, for animal i. The response time for animal i is $Y_i = \min(T_i, C_i)$ and the event indicator δ_i takes the value 1 if the event of interest occurs and 0 otherwise. A commonly used model for such data is a Cox proportional hazards model with two nested frailties $u^{(3)}_{region(i)}$ and $u^{(2)}_{herd(i)}$ acting multiplicatively on the baseline hazard (Rondeau et al., 2006) to take into account unmeasured herd and ecological region factors (the numbers in superscript parentheses represent the hierarchical levels). The conditional hazard function of the nested frailty model can be written as,

$$\lambda_i\left(t\middle|u^{(3)}_{region(i)}, u^{(2)}_{herd(i)}\right) = \lambda_0(t)u^{(3)}_{region(i)}u^{(2)}_{herd(i)}\exp(\boldsymbol{\beta}'\boldsymbol{X_i}) \qquad (4.1)$$

where $\lambda_0(.)$ is the baseline hazard, $u^{(3)}_{region(i)}$ and $u^{(2)}_{herd(i)}$ are two nested frailties following a particular probability distribution, $\boldsymbol{X_i}$ is the covariate vector for the i^{th} animal, and $\boldsymbol{\beta}$ is the corresponding vector of regression parameters. Model (4.1) can be rewritten in random effects context as,

$$\lambda_i\left(t\middle|b^{(3)}_{region(i)}, b^{(2)}_{herd(i)}\right) = \lambda_0(t)\exp\left(b^{(3)}_{region(i)} + b^{(2)}_{herd(i)} + \boldsymbol{\beta}'\boldsymbol{X_i}\right) \qquad (4.2)$$

where the frailty and random effect terms are linked by: $u = \exp(b)$.

One approach to fit model (4.2) is to utilize the relationship between the Cox model and a suitable Poisson model to translate the nested random effects Cox model into a nested random effects Poisson model (Rabe-Hesketh and Skrondal, 2012, chapter 15). As shown by Ma et al. (2003) and Feng et al. (2005), the likelihood function of Cox models with normal random effects (i.e., lognormal frailties) is proportional to the likelihood function of such random effects Poisson models. In detail, Cox models with normal random effects can be estimated as generalized linear mixed models (GLMMs) with a binary Poisson count response and a specific offset. The approach requires each observation in the data to be split into a multiple records based on the complete set of failure times in the dataset, and the offset equals the logarithm of the length of each time interval. The baseline hazard is modeled as a smooth function of time, in our case a 4^{th} order polynomial as suggested by Rabe-Hesketh and Skrondal (2012).

Using available software for GLMMs, random effects Cox models can be fitted to survival data with several hierarchical levels and more complex data structures.

4.3.2.2. Cross-classified and multiple membership frailty models

The full structure of the calf mortality data described in Section 4.2 can be taken into account through a CMM random effects Cox modeling approach. The model accounts for the cross-classified factors of veterinary clinics and ecological regions, and uses weights for the multiple membership relation of herds to the veterinary clinics so that each herd will have weights for all the veterinary clinics that the herd is serviced by. The CMM Cox model can be written as,

$$\lambda_i\left(t|b^{(3)}_{region(i)}, (b^{(3)}_j)_{j\in clinic(i)}, b^{(2)}_{herd(i)}\right)$$

$$= \lambda_0(t)\exp\left(b^{(3)}_{region(i)} + \sum_{j\in clinic(i)} w^{(3)}_{ij} b^{(3)}_j + b^{(2)}_{herd(i)} + \boldsymbol{\beta}'\boldsymbol{X_i}\right) \qquad (4.3)$$

where $b^{(2)}_{herd(i)}$ is the herd random effect, $b^{(3)}_{region(i)}$ is the ecological region random effect, and the term $\sum_{j \in clinic(i)} w^{(3)}_{ij} b^{(3)}_{j}$ involves a set of veterinary clinic random effects $b^{(3)}_{j}$ and weights $w^{(3)}_{ij}$ assigned to each herd for their veterinary clinic group membership with $\sum_{j \in clinic(i)} w_{ij} = 1$.

Assuming normal random effects, the CMM Cox model can be estimated in a Poisson modeling framework for a survival time response as described in Section 4.2.2.1 using Markov Chain Monte Carlo (MCMC) techniques and Bayesian inference.

4.3.2.3. MCMC estimation and Bayesian inference

MCMC estimation employed three chains for diagnostic purposes, 100,000 estimation samples, and a burn-in of 5,000 samples. The three chains used different initial values, were specified in Stata/MP 12.1 and run one at a time in MLwiN software version 2.25 called from within Stata using the runmlwin utility (Leckie and Charlton, 2013). The vague priors were: a uniform prior $p(\beta) \propto 1$ (flat prior) for the fixed effect parameters and a gamma $(10^{-3}, 10^{-3})$ for the inverse variances of the normal random effects. The Raftery-Lewis diagnostic (Raftery and Lewis, 1992) provided in MLwiN and the ratio rule of Monte Carlo (MC) error to the standard deviation (Lunn et al., 2013, p. 78) were used to determine the needed number of MCMC samples. The ratios of the MC error to the standard deviations for all model parameters were all less than 5%. Also the Raftery-Lewis diagnostic indicated that 100,000 samples were sufficient for estimation. Other Markov chain diagnostics, including all those given by Gelman and Rubin (1992) and implemented in R software version 2.15.3 (coda package), were carried out and found to be satisfactory.

Significance for single parameter effects in Bayesian inference was assessed using 95% credible intervals (whether or not zero lies in such intervals) or by computing a tail

probability of the posterior distribution; such probability is analogous to P-value in frequentist statistics.

4.3.3. Data analysis

4.3.3.1. Model building

Descriptive analyses were carried out for explanatory variables listed in the dataset to check distributions and invalid values, as well as to identify collinearity among variables. To facilitate the first stages of the analyses, unconditional (simple) associations between each explanatory variable and the outcome were obtained from a standard Cox regression model with the Breslow method for ties. A liberal p-value of 0.20 was chosen to determine potential important explanatory variables. Using lowess smoothing graphs, functional forms of continuous variables were evaluated by plotting the variable in question against martingale residuals, and if necessary appropriate transformation was performed or a quadratic term was added to the model.

The second step of the model building consisted in a stepwise backward selection for the standard Cox model with $P < 0.10$ as inclusion criterion since a hierarchical Cox model was impractical and very time consuming. All two-way interactions between predictors retained in the model were evaluated and tested for statistical significance; interactions that turned out significant and biologically meaningful were kept in the model. During the selection process, the non-significant predictors were rechecked for confounding and a change of 20% or more in the parameter estimate was used as a criterion for identifying confounders. The proportional hazards assumption was evaluated for model predictors individually and globally by a statistical test based on the scaled Schoenfeld residuals (Dohoo et al., 2009). To account for non-proportional hazards for some predictors, the dataset was split at events, and an interaction term between the predictor in question and the logarithm of time was added to

the model. The assumption of independent censoring was checked by sensitivity analysis comparing the change of positive and negative correlation scenarios between censoring and new mortality events. All descriptive statistics and model building were performed in Stata/MP 12.1.

4.3.3.2. Accounting for data structure

The CMM structure of the data was accounted for by including random effects for herds, veterinary clinics and ecological regions, as described in Section 4.2.2.2. The multiple membership weights of veterinary clinics servicing a given herd were computed as proportions of visits of that herd by each clinic, out of the total number of visits to the herd.

4.4. Results of calf mortality data analysis

4.4.1. Descriptive statistics

The overall mortality observed in the 174 herds was 3.8% (897/23409) with a 90% range across herds of (0.5%, 7.5%), and the percentages of calf loss occurred within the first 1 day, 3 days, 7 days, 14 days and 30 days at risk were, respectively, 19% (171/897), 29% (258/897), 40% (362/897), 55% (496/897) and 68% (614/897). The medians of event time (calf death) and censoring time were 12 and 104 days, respectively. The full list of predictor variables included in the analysis is shown in Table 1 (animal-level predictors) and Table 2 (herd-level predictors) with descriptive statistics.

Predictors selected for further consideration during the model building process ($P < 0.20$) were: whether the calf was a twin, calf gender, calving assistance, cow age, cow breed type, cow body condition at pregnancy test, within-herd calving proportion, mean 7-day temperature, shelters provided for calves separate from cows and heifers, cow-calf pairs moved to a nursery pasture within 48 hrs of birth, and whether any animals were purchased in

the month prior to or during calving. The variables of twin, surgical assistance at calving, and the mean 7-day temperature after calving were identified as time-varying effects.

4.4.2. Multivariable analysis

4.4.2.1. Model comparisons

In the Poisson modeling approach, the best model fit (i.e. the model with the smallest deviance information criteria (DIC); Spiegelhalter et al., 2002) was obtained using the logarithm of time to model both time-varying predictor effects and the baseline hazard.

To demonstrate the utility of CMM modeling for the calf mortality data, results are shown for three survival models including a standard Cox model neglecting the hierarchical structure present in the data (model 1), a Cox model with random herd effects ignoring the top hierarchical level in the data (model 2), a CMM Cox model taking into account the full hierarchical data structure (model 3). The three models were fitted to the dataset where continuous predictors were centered at the mean and the predictors twin, calving assistance and average of 7-day temperature were modeled with time-varying effects (by adding interactions with log of time). Results from the final models are tabulated in Table 4.3.

In model 2, the random herd variance parameter was estimated at 0.334 (posterior mean), with 95% credible interval (95% CI) of [0.215, 0.484]. When accounting for the full hierarchical data structure (model 3), the random herd variance estimate decreased by 19% to 0.272 [95% CI; 0.168, 0.409]. The variance for ecological regions was estimated to be about four times greater than the variance for veterinary clinics with corresponding posterior standard deviations as large as the point estimates.

Model 3 explained a greater portion of the survival outcome variation than model 2 due to handling the third hierarchical level in the dataset (the veterinary clinics and the ecological

regions). The DIC was the smallest for model 3 among the three models indicating a better model fit.

The standard Cox model (model 1) estimated with a Bayesian approach as a Poisson model gave similar estimates to those from a Cox model using a frequentist (classical) approach (results not shown). Some differences in estimates were seen compared with the random effects models (models 2 and 3). Further, the standard errors of regression coefficients from the simple Cox model ignoring the data structure were smaller than those from the CMM Cox model, especially (and as expected) for the herd-level predictors.

4.4.2.2. Interpretations of effects from model 3

In model 3, the effect of twin birth on the hazard of calf mortality depended on time and remained statistically significant until day 22 from calving. The hazard ratio (HR) for twin-birth calves relative to single-birth calves of age 1 day was estimated to be 3.80 with 95% credible interval (95% CI) of [2.70, 5.25]. Similarly, the HRs of twin-birth calves compared with single-birth calves of age 7, 22 and 60 days were estimated at 2.07 [95% CI; 1.58, 2.68], 1.45 [95% CI; 1.02, 2.01] and 1.06 [95% CI; 0.66, 1.63], respectively, suggesting that the hazard of mortality at any given time before 60 days of age was highest for twin-birth calves, and such that hazard ratios declined over time until vanishing after about two months of age. The HR for male (versus female) calves was 1.16, 16% higher hazard in males than in females at any point in time.

For calving assistance, the HR of calves that were born with a hard pull or malpresentation relative to calves born without calving assistance were 2.50 and 1.71, respectively, and thus associated with substantially higher hazard of mortality. The effect of caesarean section surgery versus unassisted varied with time: days 1, 2, 3 and day 7 had estimated HRs of 3.70 [95% CI; 1.44, 7.98], 2.29 [95% CI; 0.93, 4.82], 1.72 [95% CI; 0.65, 3.78], and 0.96 [95%

CI; 0.26, 2.52], respectively, indicating that the hazard of mortality for calves with surgical assistance at calving was higher immediately after calving and statistically significant on day 1 and then dropped down quickly.

After accounting for the other risk factors in the final model, the HRs for calves from cows aged 2, 3, 4 and greater than 10 years at calving relative to those from mature cows (5-10 years old) were estimated, respectively, to be 1.47, 1.42, 1.13 and 1.36. These results suggest that the hazard of death at any given time was greatest for calves from young (2-3 years old) and old (> 10 years old) cows, but that there was little difference in the hazard for calves of cows aged 4 years compared with calves from mature cows.

In addition, the hazard of calf mortality increased as the calving season progressed with an increasing number of calves in the herd. For instance, when the proportion of cows calving in a herd reached 0.11 and 0.91 (10% and 90% percentiles, respectively), the HRs for mortality were estimated, respectively, to be 0.90 and 1.64 compared with a proportion of 0.51 (50% percentile) indicating that the hazard of mortality increased with increasing number of births in the herd.

Modeling temperature as the mean of first 7 days post calving gave a better DIC than the temperature on day of calving. Very cold weather was associated with a high hazard of calf mortality and such hazard decreased over time. For example, when the averages of 7-day temperature post calving was 20, 10 and 5 °C below the mean (-6.42 °C), the HR of mortality relative to the mean would be, respectively, 2.53 [95% CI; 1.86, 3.36], 1.57 [95% CI; 1.36, 1.83] and 1.24 [95% CI; 1.17, 1.35] for calves of age 24 hrs; and 1.70 [95% CI; 1.39, 2.08], 1.30 [95% CI; 1.18, 1.44] and 1.13 [95% CI; 1.08, 1.20] for calves of 7 days of age; and 1.27 [95% CI; 1.01, 1.61], 1.12 [95% CI; 1.00, 1.27] and 1.06 [95% CI; 1.00, 1.13] for calves of 30 days of age, suggesting that the hazard of calf mortality was greatest if calving took place

in very cold weather. After about month of age, the time varying effect of temperature was statistically non-significant.

Three herd-level predictors related to biosecurity practices were also important predictors of calf mortality. The estimated HR was lower ($HR = 0.79$ with a probability analogous to P-value of 0.048) for calves from herds where the owner provided shelters for calves separate from cows and heifers as well as for calves from herds where cow-calf pairs were moved to a nursery pasture within 48 hrs of birth ($HR = 0.79$ with probability of 0.031). Calves from herds where any animals were purchased in the month prior to or during calving were at higher risk of death ($HR = 1.33$ with probability parallel to P-value of 0.020).

4.3.2.3. Non-significant effects

The variables cow breed type and cow body condition at pregnancy test had no effect on the hazard of calf mortality (i.e., these predictors did not contribute substantially to the model DIC) and were not included in the multivariable model. The HRs for continental and cross breeds relative to British types of breed after accounting for other risk factors were estimated, respectively, to be 1.03 [95% CI; 0.85, 1.25] and 1.13 [95% CI; 0.81, 1.58], whereas the HRs for cow body condition score at pregnancy test and pre-calving (< 5 vs. ≥ 5) were 1.05 [95% CI; 0.82, 1.33] and 0.99 [95% CI; 0.72, 1.36], respectively.

4.5. Simulation studies

Two simulation studies were conducted to evaluate the performance of the cross-classified and multiple membership Cox modeling approach discussed above. In simulation study I, the data structure and the magnitudes of variation at different levels were similar to the calf mortality dataset. In simulation study II, a more pronounced multiple membership data

structure and larger variations at different levels were considered. Both simulation studies used 200 simulated datasets.

In order to reduce the computing time of the simulations, the simulation structures were based on a subset of the real data after eliminating randomly 75% of non-cases. This reduction increased the prevalence of calf mortality to 14%. The reduced dataset had 6519 observations and the same hierarchical structure as the full data. Analysis of the reduced data showed only minor changes in model estimates compared with the results of the full data (results not shown).

4.5.1. Data structure and model parameters

Similar to the reduced real dataset, a total of 6519 animals from 174 different herds (from 3 to 111 animals per herd) were considered. In simulation study I, the data structure and multiple membership weights were the same as in the real data. In simulation study II, herds were considered to be registered in 1, 2 and 3 veterinary clinics with proportions of 52%, 25% and 23%, respectively. One dichotomous animal-level predictor was used in the two simulation models. The true values of model parameters and other features for each simulation study are presented in Table 4.4.

4.5.2. Simulating data

Using the technique of Bender et al. (2005), 200 simulated datasets for each simulation study were generated from model (4.3) using R software version 2.15.3. In each dataset, the random herds, random veterinary clinics, and random ecological regions were generated independently from a normal distribution with mean of zero and variances σ_0^2, σ_1^2, and σ_2^2, respectively. The weights in simulation study II were assigned as follows: if a herd was serviced by 3 veterinary clinics, weights for the first two clinics were randomly generated

from a uniform distribution $U(0,1)/2$ and the complement of the sum of these weights was assigned as weight for the third clinic; if a herd was serviced by 2 clinics the weight of the first clinic was randomly generated from $U(0,1)$ and the complement of that weight was the weight for the second clinic and 0 otherwise; and for herds that visited by one clinic a weight of 1 was assigned to that clinic and 0 otherwise. The fixed effect predictor was generated in each simulation from a Bernoulli distribution with a probability of 0.5.

The mortality time T_i for animal i was randomly generated from a Weibull distribution with shape parameter $P = 0.4$ and scale parameter equal to the intensity $\lambda_i(t|.)$ defined in (4.3). The time at risk C_i was randomly generated from a normal distribution with mean $\mu = 105$ and standard deviation $\sigma = 32$, censored to the interval (0.5, 180). Censoring occurred when the mortality time T_i was longer than the time at risk C_i, i.e. $Y_i = \min(T_i, C_i)$ and $\delta_i = I(T_i, C_i)$. These simulation settings led to approximately 86% censoring animals which was equivalent to the censoring rate in the reduced version of the calf mortality data.

Finally, to reduce the computing time in the simulations, the MCMC sampler was run for 55,000 iterations in each simulated model of which the initial 5,000 iterations were discarded as burn-in. The same MCMC diagnostics as described in Section 2.2.3 were carried out for selected simulated datasets and all were satisfactory.

4.5.3. Calculating summary statistics

The posterior mean, median, standard deviation, and 95% CI end points for each simulated dataset were extracted, and averages and empirical standard deviations were computed across the simulated datasets. Absolute relative bias was computed as the absolute value of the difference between the averaged estimate and the true value divided by the true value, and the mean squared error (MSE) was computed as the average of the squared differences between the estimated values and the true value over the simulated datasets.

4.5.4. Simulation results

The results of the two simulation studies are presented in Table 4.5. In simulation study I, the fixed effect β and the variance of random herd effect σ_0^2 were estimated well with relative biases not exceeding 2%. Further, the "model-based standard errors" (posterior sd) of β was on average very close to its empirical standard deviation, and the probability converges of β and σ_0^2 were somewhat over the nominal. For σ_1^2, and σ_2^2, the average posterior medians were very close to the true values, but the average posterior means were larger than the true values, with substantial relative biases of 48% and 35%, respectively. The σ_1^2 estimate showed strongly CI over-coverage, whereas the estimate of σ_2^2 had CI under-coverage. The mean squared errors were similar to the posterior-mean and posterior-median estimates of the β and σ_0^2, and smaller mean squared error for the posterior-median estimates of the σ_1^2 and σ_2^2 than for the posterior-mean estimates.

In simulation study II, the β, σ_0^2 and σ_1^2 were estimated very well based on both the posterior means and posterior medians with relative biases of at most 1.1%, 3.4% and 4%. For σ_2^2, estimation based on the posterior medians performed better than for posterior means. The CI converges of all model estimates were good except for σ_1^2 where CI under-coverage was observed.

4.6. Discussion

4.6.1. Calf mortality data

The calf mortality rate reported for these herds is slightly higher than in most previous Canadian studies with exception of one from Quebec (McDermott et al., 1991; Dutil et al., 1999; Waldner, 2001). However, our analysis included all losses from 1 hour after birth, rather than from 24 hours of age. When the calf losses after 24 hours were summarized for

the present study, the average risk of mortality was 3.1%. In an on-farm study of 7 Alberta herds over a 12-year period, Waldner (2001) and Waldner et al. (2001) reported median risks of calf mortality between 24 hours of age and weaning of 3.3% and 3.5%, similar to earlier reports from Ontario of 3.3% for first-calf heifers and 2.6% for mature cows (McDermott et al., 1991). The age distribution of calf deaths was also similar to what was expected based on other reports. The 1986 to 1987 survey by Alberta Agriculture found that 52% of deaths of calves occurred in the first 14 days compared to 55% in the current study (Mathison, 1993). The mortality rates for calves that died between one hour and 3 days of age (1.1%) and in the first 30 days (2.6%) were slightly higher than the 0.7% and 1.6% reported from a 2010 mail survey of 303 herds from western Canada (Waldner et al., 2013).

The large observational data set and time-to-event analyses provided us with a unique opportunity for an intensive assessment of risk factors for calf mortality reported in previous papers, as well as an opportunity to explore new environmental and herd management variables. For example, an association between twin birth and average calf mortality from 12hr to 45 days was reported in a previous study of 10 herds in Colorado (Wittum et al., 1994). Gregory et al. (1996) also reported higher survival rates for singles as compared to twins from one research herd at 72 hours and 150 days when there was no requirement for assistance. Our study found that while the death rate is highest for twins in the perinatal period, there is a significant increased risk of loss in privately owned commercial calves up to 22 days of age after accounting for other risk factors.

The increased risk of mortality for male calves remained constant throughout the observation period similar to what was observed using unconditional analysis by Patterson et al. (1987). Azzam et al. (1993) also reported an increase in mortality for bull calves after accounting for dystocia and the relative calf size. The paper is different in that the authors used data from a research centre and included all calves that were alive at the start of calving.

The higher mortality rate for calves classified as having a hard pull or malpresentation at birth did not decrease during the study period. In contrast, the increased death rate for calves born by caesarean section was only significant for day 1 and was only elevated for the first week. While other studies have identified dystocia as a risk factor for perinatal calf mortality (Wittum et al., 1994; Ganaba et al., 1995), only one study in a single research herd specifically explored the longer term effects on calf survival using individual data (Gregory et al., 1996). Dutil et al. (1999) reported a weak association between herd dystocia rates and preweaning mortality in 148 Quebec herds, but did not account for confounding by individual factors such as parity.

After accounting for all other known risk factors, the only important cow attribute was age. Previous studies have identified increased postnatal calf loss from heifers in addition to the well-established increased risk of loss in heifers' calves that died at or very near birth (Wittum et al., 1994). Our study is unique in that higher risks of postnatal calf mortality were also identified for cows having their second calf and cows ≥ 10 years old. The increased risk of postnatal calf mortality for each cow age group was consistent throughout the follow up period. Others looking at cow age either had a smaller sample size and did not see a difference (Wittum et al., 1994) or looked at all mature cows together and did not differentiate older cows (Azzam et al., 1993).

After accounting for cow age and assistance at calving there was no difference in calf survival across the range of observed body condition scores. While others have documented an association between poor nutrition in the last trimester and calf mortality due to scours (Corah et al., 1975), $< 5\%$ of cows in this cohort were thin at calving, thus providing very little power to examine this hypothesis.

While a number of authors have suggested that as the calving grounds become more contaminated the risks of calf morbidity and mortality increase, there have been no previous studies that test this hypothesis across a large number of herds. Schumann et al. (1990) reported that as the proportion of the nursing area that was poorly drained, wet and muddy increased the odds of mortality from diarrhea also increased. In this study, we looked at the contextual effect of when the calf was born in relation to the other calves in the herd. The idea was simply that calves that are born later in their cohort are potentially exposed to a greater build-up of pathogens. In this study, there was a substantial increase in the mortality rate for calves born after the half-way point in each herd. Clement et al. (1995) had previously documented increased odds of developing diarrhea in calves born after the median calving date. They hypothesized that the numbers of diarrhea-causing pathogens increased during the calving season.

While some previous studies have used postmortem findings (Bellows et al., 1987) and owner reported cause of loss to document the importance of calf deaths due to cold weather (Wittum et al., 1993), only one other study has actually looked at meteorological conditions (Azzam et al., 1993). This study like ours found an increased calf mortality rate for calves born under cold conditions. Because the other study was limited to a single research herd, they also had access to local precipitation data which were not consistently available in the present analysis. Azzam et al. (1993) used logistic regression to examine effect of temperature on the day of birth on total risk of calf loss from birth to weaning. However, this study did not account for repeated measures in the analysis or consider whether the effect of meteorological conditions at birth changed with calf age. We used the average temperature for the first week after birth and demonstrated that for calves born in very cold weather (< -10 °C), the associated hazard extended through the first month of life.

Because of the relatively large number of herds compared to previous studies we were also able to evaluate a number of common management and biosecurity practices. In our study, herd owners that moved calves out of the calving area and to a nursery pasture within 48 hours had lower calf losses. This practice removes cow-calf pairs from the contaminated environment and prevents crowding in the calving area by dispersing newborn calves soon after birth (Radostits and Acres, 1980). The use of calf shelters which are not accessible to cows and heifers (Radostits and Acres, 1980; Olson, 1986) was also associated with decreased calf mortality. Schumann et al. (1990) reported that increasing the nursery shelter area helped to protect against calf diarrhea, but did not differentiate between shelters accessible to both cows and calves and shelters accessible to just calves.

Finally, herds where any cattle were purchased in the month before or during calving had higher calf mortality rates than those that did not. Schumann et al. (1990) reported a similar unconditional association between replacing dead calves with purchased calves less than one month of age and higher odds of calf mortality. We did not see an increased rate of loss specifically associated with the purchase of foster calves; however, this practice was uncommon in the current study.

The analysis of calf mortality data demonstrated a larger variation between herds than between both veterinary clinics and ecological regions, and a clear improvement in model fit after accounting for the variation between veterinary clinics and ecological regions.

4.6.2. Simulations

In the setting similar to the real data (study I), the results indicated that the proposed model performed well in estimating most of the model parameters if posterior medians were used for the inference and overestimated the between-clinic and between-ecoregion variances

when the inference was based on posterior means. The simulation study therefore supported our findings of relatively small variance components for veterinary clinics and regions in the real data. In addition, simulation study II showed that the estimation of between-clinic variance was improved in a more pronounced multiple membership structure and with larger variance components. Both simulation studies demonstrated difficulties with estimation of the between-ecoregion variance and its standard error, and this can probably be attributed to the small number of ecologic regions.

We finally note that the performance of the proposed model and estimation can depend on many parameters, for instance, the censoring rate, the shape of baseline hazard, the number of clusters, the cluster size, and the magnitude of heterogeneity. A detailed exploration of how such parameters might affect performance is beyond the scope of the present study, but could be a topic for future investigation.

Acknowledgements

The data in this paper were collected as part of the field research activities for the Western Canada Study of the Animal Health Effects Associated with Exposure to Emissions from Oil and Natural Gas Field Facilities. Funding support was provided through the Western Interprovincial Scientific Studies Association. The authors appreciate the dedication of all of the veterinarians and producers who contributed to this project.

References

Azzam S, Kinder J, Nielsen M, Werth L, Gregory K, Cundiff L, Koch R. (1993). Environmental effects on neonatal mortality of beef calves. *Journal of Animal Science*, **71**: 282-290.

Bellows R, Patterson D, Burfening P, Phelps D. (1987). Occurrence of neonatal and postnatal mortality in range beef cattle. I. Factors contributing to calf death. *Theriogenology*, **28**: 573-586.

Bender R, Augustin T, Blettner M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, **24**: 1712-1723.

Browne W, Goldstein H, Rasbash H. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, **1**: 103-124.

Chung H, Beretvas N. (2012). The impact of ignoring multiple membership data structures in multilevel models. *British Journal of Mathematical and Statistical Psychology*, **65**: 185-200.

Clement J, King M, Salmon M, Wittum T, Casper H, Odde K. (1995). Use of epidemiologic principles to identify risk factors associated with the development of diarrhea in calves in five beef herds. *Journal of the American Veterinary Medical Association*, **207**: 1334-1338.

Corah L, Dunn T, Kaltenback C. (1975). Influence of prepartum nutrition on the reproductive performance of beef females and the performance of their progeny. *Journal of Animal Science*, **41**: 819-824.

Dohoo I, Martin W, Stryhn H. (2009). *Veterinary Epidemiologic Research*. 2nd edition, Charlottetown, VER - Inc.

Dutil L, Fecteau G, Bouchard E, Dutremblay D, Pare J. (1999). A questionnaire on the health, management, and performance of cow-calf herds in Quebec. *The Canadian Veterinary Journal*, **40**: 649-656.

Feng S, Wolfe R, Port F. (2005). Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *Journal of the American Statistical Association*, **100**: 728-735.

Fielding A, Goldstein H. (2006). Cross-classified and multiple membership structures in multilevel models: An introduction and review (Research Report Number 791). Birmingham, UK: University of Birmingham, Department of Education and Skills.

Ganaba R, Bigra-Poulin M, Belanger D, Couture Y. (1995). Description of cow-calf productivity in Northwestern Quebec and path models for calf mortality and growth. *Preventive Veterinary Medicine*, **24**: 31-42.

Gelman A, Rubin D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **4**: 457-511.

Goldstein H, Browne W, Rasbash J. (2002). Tutorial in biostatistics: Multilevel modelling of medical data. *Statistics in Medicine*, **21**: 3291-3315.

Goldstein H, Burgess S, McConnell B. (2007). Modelling the effect of pupil mobility on school differences in educational achievement. *Journal of the Royal Statistical Society: Series A*, **170**: 941-954.

Gregory K, Echternkamp S, Cundiff L. (1996). Effects of twinning on dystocia, calf survival, calf growth, carcass traits, and cow productivity. *Journal of Animal Science*, **74**: 1223-1233.

Leckie G, Charlton C. (2013). runmlwin: A program to run the MLwiN multilevel modelling software from within Stata. *Journal of Statistical Software*, Volume 52, Issue 11.

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. (2013). *The BUGS Book: A Practical Introduction to Bayesian Analysis.* Boca Raton: Taylor & Francis Group, LLC.

Luo W, Kwok O. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*, **44**: 182-212.

Ma R, Krewski D, Burnett R. (2003). Random effects Cox models: A Poisson modeling approach. *Biometrika*, **90**: 157-169.

Masaoud E, Stryhn H, Whyte S, Browne W. (2011). Statistical modeling of neighbour treatment effects in aquaculture clinical trials. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**: 202-220.

Mathison G. (1993). The beef industry. In: Martin J, Hudson R I, Young BAE, eds. Animal Production in Canada. Edmonton, Alberta: University of Alberta, pp. 35–74.

McDermott J, Alves D, Anderson N, Martin S. (1991). Measures of herd health and productivity in Ontario cow–calf herds. *The Canadian Veterinary Journal*, **32**: 413-420.

Meyers J, Beretvas N. (2006). The impact of inappropriate modelling of cross-classified data structures. *Multivariate Behavioral Research*, **41**: 473-497.

Patterson D, Bellows R, Burfening P, Carr J. (1987). Occurrence of neonatal and postnatal mortality in range beef cattle. I. Calf loss incidence from birth to weaning, backward and breech presentations and effects of calf loss on subsequent pregnancy rate of dams. *Theriogenology*, **28**: 557-571.

Olson D. (1986). Field studies of protective shelters for beef calves. *Bovine Practitioner*, **21**: 19-22.

Rabe-Hesketh S, Skrondal A. (2012). *Multilevel and Longitudinal Modeling using Stata*. *Volume II: Categorical Responses, Counts, and Survival*, 3rd ed, Stata Press.

Radostits O, Acres S. (1980). The prevention and control of epidemics of acute undifferentiated diarrhea of beef calves in Western Canada. *The Canadian Veterinary Journal*, **21**: 243-249.

Raftery A, Lewis S. (1992). How many iterations in the Gibbs Sampler? In Bayesian Statistics 4. Bernardo, J., Berger, J., Dawid, A., Smith, A. (Eds.), Oxford University Press, pp. 763-773.

Rondeau V, Filleul L, Joly P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine,* **25**: 4036-4052.

Schumann F, Townsend H, Naylor J. (1990). Risk factors for mortality from diarrhoea in beef calves in Alberta. *Canadian Journal of Veterinary Research*, **54**: 336-372.

Spiegelhalter D, Best N, Carlin B, van der Linde A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B,* **64**: 583-640.

Waldner C. (2001). Monitoring beef cattle productivity as a measure of environmental health. *Environmental Research Section A*, **86**: 94-106.

Waldner C, Ribble C, Janzen E, Campbell J. (2001). Associations between total sulphation, hydrogen sulphide deposition, and beef cattle breeding outcomes in Western Canada. *Preventive Veterinary Medicine*, **50**: 19-33.

Waldner C. (2008). Western Canada study of animal health effects associated with exposure to emissions from oil and natural gas field facilities. Study design and data collection I. Herd performance records and management. *Archives of Environmental & Occupational Health No 4*, **63**: 167-184.

Waldner C, Jelinski M, McIntyre-Zimmer K. (2013). Survey of western Canadian beef producers regarding calf-hood diseases, management practices, and veterinary service usage. *The Canadian Veterinary Journal*, **65**: 559-420.

Wittum T, Salman M, Odde K, Mortimer R, King M. (1993). Causes and costs of calf mortality in Colorado beef herds participating in the National Animal Health Monitoring System. *Journal of the American Veterinary Medical Association*, **203**: 232-236.

Wittum T, Salman M, King M, Mortimer R, Odde K, Morris D. (1994). Individual animal and maternal risk factors for morbidity and mortality of neonatal beef calves in Colorado, USA. *Preventive Veterinary Medicine*, **19**: 1-13.

Table 4.1. Descriptive statistics for animal-level predictors in the calf mortality dataset (23,409 calves from 174 herds) including the proportion of calves and the probability of calf mortality for each categorical variable, as well as the mean and standard deviation (sd) for continuous variables.

Predictor		Proportion	Mortality probability
Calf gender	(female)	0.48	0.035
	(male)	0.52	0.041
Twin	(single)	0.96	0.037
	(twin)	0.04	0.069
Cow age at calving	(≤ 2 years old)	0.18	0.049
	(3 years old)	0.17	0.046
	(4 years old)	0.12	0.036
	(5 to 10 years old)	0.45	0.032
	(>10 years old)	0.08	0.044
Cow breed type	(British)	0.43	0.038
	(continental)	0.49	0.037
	(cross)	0.08	0.046
Calving assistance	(unassisted)	0.91	0.037
	(easy pull)	0.05	0.040
	(hard pull)	0.02	0.102
	(malpresentation)	0.01	0.071
	(surgery)	0.01	0.067
Cow body condition at pregnancy test	(BCS<5)	0.09	0.046
	(BCS≥5)	0.91	0.038
Cow pre-calving body condition score	(BCS<5)	0.04	0.047
	(BCS≥5)	0.96	0.038
Cow problems following calving [a]	(yes)	0.01	0.052
	(no)	0.99	0.038

Predictor	Mean (sd) (cases)	Mean (sd) (non-cases)
Average temperature (Celsius) for the first 7 days post calving	-7.353 (8.198)	-6.362 (8.483)
Within-herd calving proportion at calving	0.516 (0.291)	0.514 (0.286)

[a] Cow problems including retained placentas, uterine prolapses, and metritis.

[b] Computed as a number of new calves at a particular calving day divided by the total number of calves in a herd.

Table 4.2. Descriptive statistics for herd-level predictors in the calf mortality dataset (174 herds) including the proportion of herds and mortality probability for each categorical variable.

Predictor		Proportion	Average of mortality[a]
Cows due to calve and cows that have calved are together	(yes)	0.37	0.037
	(no)	0.63	0.039
Heifers due to calve before rest of the cow herd	(yes)	0.29	0.042
	(no)	0.71	0.037
Provide shelters for calves separate from cows and heifers	(yes)	0.79	0.038
	(no)	0.21	0.041
Move cow-calf pairs to a nursery pasture within 48 hrs of birth	(yes)	0.70	0.038
	(no)	0.30	0.040
Buy foster calves (Holstein bull calves)	(yes)	0.10	0.038
	(no)	0.90	0.038
Were any animals purchased in the month prior to or during calving	(yes)	0.76	0.039
	(no)	0.24	0.035
Cows vaccinated for *E. coli* prior to calving	(yes)	0.49	0.040
	(no)	0.51	0.037
Heifers vaccinated for *E. coli* prior to calving	(yes)	0.53	0.038
	(no)	0.47	0.038
Cows vaccinated for rota/corona virus prior to calving	(yes)	0.50	0.038
	(no)	0.50	0.039
Heifers vaccinated for rota/corona virus prior to calving	(yes)	0.53	0.037
	(no)	0.47	0.040

[a]Average of within herd mortality.

Table 4.3. Parameter estimates for the analysis of calf mortality dataset: the mean, median, and standard deviation (sd) of posterior distribution from a standard Cox model (model 1), a Cox model with random herd effect (model 2), and a CMM Cox model (model 3).

	Model 1			Model 2			Model 3		
Predictor/Parameter[a]	Mean	Median	sd	Mean	Median	sd	Mean	Median	sd
Twin									
Twin vs. single	0.145	0.149	0.199	0.205	0.208	0.197	0.213	0.218	0.202
Twin × T[b]	-0.338	-0.338	0.077	-0.331	-0.331	0.075	-0.331	-0.330	0.077
Calf gender									
Male vs. female	0.154	0.154	0.068	0.151	0.152	0.068	0.150	0.150	0.068
Calving assistance									
Easy pull vs. unassisted	-0.105	-0.103	0.150	-0.050	-0.048	0.151	-0.049	-0.046	0.150
Hard pull vs. unassisted	0.872	0.877	0.169	0.921	0.924	0.173	0.917	0.920	0.174
Malpresentation vs. unassisted	0.532	0.541	0.253	0.535	0.543	0.251	0.537	0.546	0.254
Surgery vs. unassisted	-1.304	-1.202	0.980	-1.243	-1.131	0.976	-1.198	-1.090	0.960
Surgery × T	-0.737	-0.720	0.300	-0.743	-0.725	0.298	-0.739	-0.721	0.294
Cow age									
≤ 2years old vs. 5-10 years old	0.401	0.402	0.095	0.382	0.381	0.096	0.384	0.384	0.096
3 years old vs. 5-10 years old	0.340	0.341	0.095	0.342	0.342	0.095	0.349	0.350	0.096
4 years old vs. 5-10 years old	0.099	0.100	0.113	0.121	0.121	0.112	0.124	0.126	0.113
>10 years vs. 5-10 years old	0.332	0.333	0.126	0.308	0.309	0.127	0.305	0.307	0.127
Within-herd calving prop.									
Linear	0.631	0.631	0.128	0.759	0.758	0.137	0.767	0.766	0.136
Quadratic	1.041	1.040	0.467	1.233	1.232	0.473	1.243	1.243	0.474
Mean 7-day temperature[c] **(°C)**	-0.030	-0.030	0.006	-0.090	-0.090	0.006	-0.100	-0.100	0.006
Mean 7-day temperature[c] × T	0.100	0.100	0.003	0.110	0.110	0.003	0.110	0.110	0.003
Provide shelters [d]									
Yes vs. no	-0.309	-0.310	0.083	-0.325	-0.325	0.144	-0.240	-0.240	0.145
Move calf pairs [e]									
Yes vs. no	-0.306	-0.307	0.074	-0.244	-0.245	0.125	-0.237	-0.237	0.126
Animals purchased [f]									
Yes vs. no	0.258	0.257	0.085	0.243	0.242	0.145	0.285	0.283	0.141
Herd variance				0.334	0.327	0.070	0.272	0.267	0.062
Veterinary clinic variance							0.024	0.012	0.030
Ecological region variance							0.099	0.073	0.100
DIC		14724.2			14566.1			14558.5	

[a] Coefficients for the 4th order polynomial of log(time) represents the baseline hazard not shown.

[b] T is a standardized log time (log time-mean/sd) or ($T = [\log(time\ in\ days) - 3.604]/1.062$).

[c] Coefficients ×10.

[d] Provide shelters for calves separate from cows and heifers.

[e] Move cow-calf pairs to a nursery pasture within 48 hrs of birth.

[f] Animal purchased in the month prior to or during calving.

Table 4.4. Model parameters and proportions of herds in a multiple membership with veterinary clinics used for the two simulation studies.

Feature	**Simulation study I**	**Simulation study II**
Fixed effect (β)	0.150	1.000
Herd variance (σ_0^2)	0.300	0.500
Veterinary clinic variance (σ_1^2)	0.025	0.500
Ecological region variance (σ_2^2)	0.100	0.500
Baseline hazard parameters* (p, λ_0)	(0.4, 0.019)	(0.4, 0.008)
Herds in a multiple membership	8%	52%

*Weibull distribution: p = shape, λ_0 = scale.

Table 4.5. Simulation study results: average of the estimates (posterior mean and posterior median) with empirical standard deviations (Esd), 95% end point confidence intervals (95% CI), and of posterior standard deviation (sd) over 200 simulated data sets, as well as probability coverage, absolute relative bias and mean squared error (MSE) for posterior mean-based and median-based estimates.

Model	**True**	**Estimate**			**95% CI**	**Prob.**	**Abs. relative bias**		**MSE (× 100)**	
param.	**value**	**mean (Esd)**	**median (Esd)**	**sd**	**end points**	**cover.**	**mean**	**median**	**mean**	**median**
Simulation study I										
β	**0.150**	.153 (.061)	.153 (.061)	.067	(.022, .284)	97%	.020	.020	0.370	0.371
σ_0^2	**0.300**	.299 (.064)	.294 (.063)	.066	(.187, .443)	97%	.003	.020	0.407	0.403
σ_1^2	**0.025**	.037 (.032)	.027 (.031)	.036	(.002, .131)	99%	.480	.080	0.115	0.094
σ_2^2	**0.100**	.135 (.099)	.100 (.080)	.127	(.021, .454)	92%	.350	.000	1.088	0.641
Simulation study II										
β	**1.000**	.990 (.074)	.989 (.074)	.075	(.844, 1.137)	95%	.010	.011	0.562	0.561
σ_0^2	**0.500**	.517 (.120)	.504 (.118)	.121	(.317, .489)	95%	.034	.008	1.453	1.381
σ_1^2	**0.500**	.509 (.216)	.480 (.210)	.205	(.194, .989)	91%	.018	.040	4.669	4.411
σ_2^2	**0.500**	.645 (.362)	.510 (.295)	.516	(.169, 1.930)	95%	.290	.020	15.17	8.656

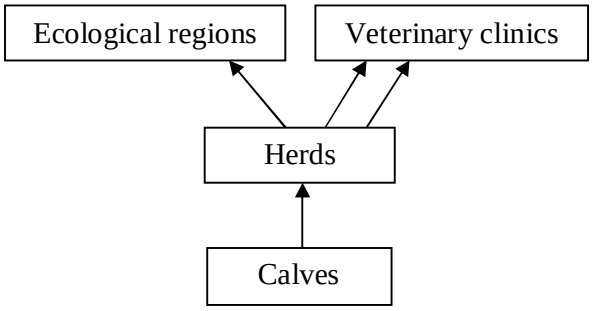

Figure 4.1. Calf mortality data structure: calves (level 1) nested in herds (level 2), herds in a multiple membership to veterinary clinics and cross-classified by ecological regions and veterinary clinics (level 3).

Chapter 5

A simulation-based assessment of misspecifying the random effects distribution in mixed-effects Cox models

The work of this chapter is intended to be submitted for publication as: Elghafghuf A, Stryhn H. (2014). A simulation-based assessment of misspecifying the random effects distribution in mixed-effects Cox models. *Computational Statistics & Data Analysis*. In preparation.

5.1. Abstract

Mixed-effects Cox models can be fit as Poisson generalized linear mixed models (GLMMs) after transforming time-to-event data to the Poisson GLMM framework. Estimation in these approximating models is based on Poisson maximum likelihood theory, assuming a specific distribution for random effects. However, the validity of the random-effects distribution assumption is often difficult to verify. In this study we assess, through simulations, the robustness of Poisson maximum likelihood estimation for a Cox model with normal random effects under misspecification of the random effects distribution. The impact of misspecifying the distribution of random effects is studied in shared frailty, random slope, and nested frailty Cox models. Factors such as the magnitude of the random effect variances, censoring rate, group size, and number of groups were accounted for in the assessment. In the simulations, the Poisson modeling approach produced robust estimates under misspecification of the random-effects distribution for fixed effects at different hierarchical levels. Non-robust estimation of variance components was observed only when the magnitude of heterogeneity, event rate, number of groups, and group size was large.

5.2. Introduction

Data in medical research are very often clustered in groups, such as health centres in a human medicine or farms in veterinary medicine studies. When the outcome of interest is time-to-event, proportional hazards models with random effects, also referred to as frailty models (Therneau and Grambsch, 2000; Duchateau and Janssen, 2008; Wienke, 2010; Hanagal, 2011), are the most common choice for modeling these type of data because these models account for the heterogeneity caused by unmeasured factors due to clustering. In frailty models, the standard assumption for the random effects (on the natural log scale) is that they follow a certain probability distribution, and the popular distribution choices are zero-mean

log-gamma and normal distributions. Other choices for frailty distributions, such as the inverse Gaussian, positive stable (Hougaard, 1995), power variance function (Aalen, 1988), and compound Poisson (Aalen, 1992) have been used in the literature.

Because the random effects are unobserved entities, it is important to study the robustness of the estimation approach against misspecification of the random effects (frailty) distribution. Previous work has been carried out to assess the robustness properties of estimators when the random effects distribution is misspecified. For instance, Sastry (1997) used the EM algorithm for gamma nested frailty model estimation, and found that the choice of frailty distribution only mattered when the frailty variance was large. Ferreira and Garcia (2001) showed that a gamma shared frailty model underestimated the between-group variance when the true model was a log-normal frailty model; they used a partial marginal likelihood (Nielsen et al., 1992) with EM algorithm for estimation. Using a penalized partial likelihood, Glidden and Vittinghoff (2004) examined the performance of the gamma shared frailty model with misspecified frailty density. For inverse Gaussian and positive stable frailty distributions, they found that the misspecified model produced similar estimates to those based on the correct model even for a small number of groups but with large group sizes. Further, based on a gamma shared frailty model and penalized partial likelihood estimation, Duchateau and Janssen (2008) pointed out that the robustness of variance component estimators is an issue when the frailty variance is large. Cortiñas et al. (2007), on the other hand, tested the estimators of the REML estimation method (McGilchrist and Aisbett, 1991), the penalized partial likelihood (Ripatti and Palmgren, 2000), the Bayesian approach (Legrand et al., 2005), and the EM algorithm with Laplace approximation (Cortiñas and Burzykowski, 2005) under misspecification of the random-effects distribution based on a Cox model with independent random clusters and random slope effects. They found that the bias

almost doubled under model misspecification relative to a model with correctly specified random-effects distribution.

The primary aim of this study was to examine the Poisson maximum likelihood estimation approach (Ma et al., 2003; Feng et al., 2005) for estimating Cox models with normal random effects against misspecification of the random effects density. Simulations based on the three different models: a Cox model with random group effect, a Cox model with random group effect and random coefficient, and a Cox model with two nested random effects were performed where the distribution(s) of the random effects were known. Factors such as magnitude of the variability of random effects, censoring rate, predictor type, group size and number of groups were controlled. A secondary aim of the study was to add information about the performance of the Poisson modeling approach for a correctly specified model.

5.3. Mixed-effects Cox models

5.3.1. Cox model with group random effect

In the following, we consider time-to-event data from a total of N individuals clustered by G groups, e.g. health centers or farms. For individual j ($j = 1, \dots, n_i$) from group i ($i = 1, \dots, G$), let T_{ij} and C_{ij} denote the event and right censoring times, respectively, where C_{ij} is independent of T_{ij}. The observed times Y_{ij} are $Y_{ij} = \min(T_{ij}, C_{ij})$, and $\delta_{ij} = I_{\{T_{ij} \leq C_{ij}\}}$ is the event indicator. For each individual, the within-group x_{1ij} and the between-group x_{2i} predictors are observed. The conditional hazard function for a Cox model with group random effect is given by

$$\lambda_{ij}(t|b_i) = \lambda_0(t)\exp\left(b_i + \beta_1 x_{1ij} + \beta_2 x_{2i}\right) \quad (5.1)$$

where $\lambda_0(t)$ is a baseline hazard, β_1 and β_2 are fixed effects coefficients corresponding to the predictors x_{1ij} and x_{2i}, and b_i is the random effect associated with group i. Alternatively, model (5.1) can be rewritten as

$$\lambda_{ij}(t|u_i) = \lambda_0(t)\, u_i \exp\left(\beta_1 x_{1ij} + \beta_2 x_{2i}\right) \tag{5.2}$$

where $u_i = e^{b_i}$ is a frailty term. Model (5.2) is known as a shared frailty model.

5.3.2. Cox model with random group and random slope effects

In many instances covariate effects may vary between groups or clusters. For example, the effect of treatment may change over trials in a meta-analysis of multicenter studies (Duchateau and Janssen, 2008; Rondeau et al., 2008) or across farms in epidemiological investigations of veterinary medicine (Stryhn and Christensen, 2013). To account for heterogeneity in the baseline hazard and predictor effects between groups, a Cox model with two random effects at the group level can be applied. Using the notation of Section 5.3.1, the hazard function for a Cox model with random intercept b_{i0} and random coefficient b_{i1} (for the predictor X_1) takes the form,

$$\lambda_{ij}(t|b_{i0}, b_{i1}) = \lambda_0(t) \exp\left(b_{i0} + b_{i1}x_{1ij} + \beta_1 x_{1ij} + \beta_2 x_{2i}\right) \tag{5.3}$$

where $\lambda_0(t)$ is the baseline hazard function, and β_1 and β_2 are regression parameters associated with the predictors X_1 and X_2.

5.3.3. Cox model with nested random effects

When survival data are clustered at several hierarchical levels (e.g. in veterinary medicine, animals from different farms and these farms located in different geographic regions), a Cox model with nested random effects (nested frailties) is used to estimate possible variation at the different hierarchical levels (Rondeau et al., 2006; Duchateau and Janssen, 2008).

Assume we have G groups and there are n_i subgroups of individuals within the i^{th} group. With x_{1ijk}, x_{2ij} and x_{3i} be the observed explanatory variables at the different hierarchical levels, the hazard function for a Cox model with nested random effects is

$$\lambda_{ijk}\left(t\middle|b_i, b_{ij}\right) = \lambda_0(t)\exp\left(b_i + b_{ij} + \beta_1 x_{1ijk} + \beta_2 x_{2ij} + \beta_3 x_{3i}\right) \quad (5.4)$$

where $\lambda_0(t)$ is the baseline hazard function, X_1, X_2 and X_3 are the explanatory variables with corresponding regression parameters β_1, β_2, and β_3, respectively. The nested terms b_i and b_{ij} are the i^{th} group and the ij^{th} subgroup random effects, respectively.

In all of the aforementioned models, the random effects b (or the frailties u) are assumed to follow a certain probability distribution.

5.4. Frailty and random effect distributions

Many probability distributions for random effects (or frailty) have been suggested in the literature. Most of the existing arguments related to the choice of frailty distribution are mathematically based. In practice, the widely used distributions are a zero-mean normal distribution for random effects (log-normal with mean 1 for frailty) and a gamma distribution with mean 1 for frailty (zero-mean log-gamma distribution for random effects). We focus on the choice of normal distribution for random effects, since it is implemented in popular statistical software packages such as Stata, SAS and R. Other distributions such as inverse Gaussian, positive stable, power variance function, and compound Poisson have been applied (Duchateau and Janssen, 2008).

To illustrate the relation between heterogeneity parameters in the two common choices of random effects and frailty distributions, note that when the random effect $b = \log u$ follows a normal distribution with zero mean and variance of σ^2, the frailty $u = e^b$ follows a log-normal distribution with mean of $\exp(\sigma^2/2)$ and variance of $\exp(\sigma^2)(\exp(\sigma^2) - 1)$.

Similarly, when the frailty u has a gamma distribution with equal shape ν and inverse scale η of $1/\theta$ (corresponding to frailty variance of θ), the random effect b has a log-gamma distribution with mean of $\Psi(1/\theta) + \log(\theta)$ and variance of $\Psi'(1/\theta)$, where $\Psi(.)$ and $\Psi'(.)$ are the digamma and trigamma functions, respectively.

In a Cox model with correlated random group and random slope effects, the random effects b_{i0} and b_{i1} are usually assumed to follow a zero-mean normal distribution since the normal distribution is more flexible than the gamma (or log-gamma) distribution for creating correlation. When the random effects b_{i0} and b_{i1} follow the bivariate normal distribution $N(0,0,\sigma_0^2,\sigma_1^2,\sigma_{01})$, the frailties u_{i0} and u_{i1} follow a bivariate log-normal distribution $LN(\mu_0,\mu_1,\theta_0,\theta_1,\theta_{01})$ with $\mu_p = \exp(\sigma_p^2/2)$, $\theta_p = \exp(\sigma_p^2)(\exp(\sigma_p^2) - 1), p = 0,1$ and covariance $\theta_{01} = \exp[(\sigma_0^2 + \sigma_1^2)/2][\exp(\sigma_{01}) - 1]$. The latter formula follows directly from the moment-generating function of a bivariate normal distribution.

This study focuses on two non-normal distributions for random effects. First, a log-gamma distribution for random effects (equivalent to gamma frailties) since the log-gamma is a non-symmetrical, left-skewed distribution (Figure 5.1). The density function of a variable b following a log-gamma distribution with shape parameter ν and inverse scale η is given by

$$f_B(b) = \frac{\eta^\nu}{\Gamma(\nu)}\exp(\nu b - \eta\exp(b)) \; ; \; -\infty < b < \infty \qquad (5.5)$$

Second, a Laplace distribution for random effects was chosen because of its higher peak and heavier tails compared with the normal distribution (Figure 5.2), as well as its flexibility for generating correlation between random effects in multivariate random effects models. The density function of a Laplace distribution with mean zero and scale parameter τ is given by

$$f_B(b) = \frac{1}{2\tau}\exp\left(\frac{|b|}{\tau}\right) \; ; \; -\infty < b < \infty \qquad (5.6)$$

5.5. Estimation approach

One approach to estimate Cox proportional hazards models is to transform the time-to-event data to the more flexible GLM framework using the equivalence of a Cox proportional hazards model with a Poisson regression model. This equivalence can be carried over to the Cox proportional hazards model with random effects (Ma et al., 2003; Feng et al., 2005; Feng et al., 2009; Rabe-Hesketh and Skrondal, 2012). The existence of statistical theory and software for GLMMs provides many opportunities to apply random-effects Cox models to complex survival data and allows for sophisticated variance structures. By using, for example, the implementation in Stata for Poisson GLMMs, adaptive Gaussian quadrature can be applied for accurate estimation.

To fit a Cox model with random effects as a Poisson GLMM model, the duration of follow-up has to be divided into increments, by splitting the data at event times, to allow for a nonparametric modeling approach to the baseline hazard. As pointed out in Feng et al. (2005), under a non-informative and independent censoring assumption, the conditional likelihood of a random-effects Cox model is proportional to the conditional likelihood of a random-effects Poisson model with the log of the interval length at risk as an offset term. The baseline hazard can be fit as a set of dummy variables representing the individual time points or as a smooth function of time by adding polynomial terms of time to the model until an additional term does not substantially improve the model fit.

5.6. Simulation studies

We study the robustness of the Poisson modeling approach for estimating Cox models with normally distributed random effects against misspecification of the random-effects (frailty) distribution based on different random-effects distributions. This aims to examine the consequences of assuming normality of random effects to distributions that are clearly non-

normal. For each simulation model, we compare the results obtained from a model with misspecified random-effects distribution to those obtained from a model with a correctly specified random-effects distribution. The robustness of the estimates is investigated for variable numbers and sizes of groups and subgroups. Further, the impact of the amount of censoring and the magnitude of the variance components are studied.

5.6.1. Description of the simulation studies

Three simulation studies were performed to study the impact of misspecifying the random effects distribution on parameter estimates obtained using the Poisson modeling approach. The structure of simulated data was built around the calf mortality data described in Waldner (2008) where calves clustered by farms (subgroups) and these farms was located in different ecological regions (groups). During a study period of 180 days, calves were followed from birth until death or censoring. Analysis of calf mortality data was reported in Chapter 4. In our simulations, we varied the number of groups and subgroups and used different sample sizes per group and subgroup. For simplicity, balanced groups and subgroups were assumed in all the simulation studies. Further, a Weibull baseline hazard with shape parameter α and scale γ was used in all simulations.

5.6.1.1. Simulation study I

Three different simulation scenarios were considered in this initial study. In each scenario, data were generated from model (5.1) assuming zero-mean normal, log-gamma, or Laplace distributions for the random effects $b_i, i = 1, \ldots, G$. In each scenario, either 20 and 70 groups with 10, 40, or 100 individuals per group were used. Because the magnitudes of heterogeneity and censoring rate are often large in animal-health data from observational studies, two degrees of censoring, for simplicity termed heavy (around 85%) and moderate (around 50%), and two different levels of between-group variance, $\sigma^2 = 0.2$ and $\sigma^2 = 0.5$,

were chosen to mirror commonly encountered situations. The true values of model parameters, including the fixed effects and variance components and the Weibull baseline hazard parameters for each scenario, are shown in Table 5.1.

5.6.1.2. Simulation study II

We considered two scenarios for this simulation study. In the two scenarios, data were simulated from model (5.3). The random effects b_{i0} and $b_{i1}, i = 1, \dots, G$, were correlated and assumed to follow a zero-mean bivariate normal distribution in the first scenario, and a bivariate Laplace distribution with zero means in the second scenario. The same number of groups and individuals per group as in simulation study I were used. Similar to simulation study I, two different sizes of the variance components were studied in high and moderate censoring rate settings. The model parameters for each simulation setting were set as described in Table 5.1.

5.6.1.3. Simulation study III

Data were generated from model (5.4) in two different scenarios. In the first, the nested random effects b_i and $b_{ij}, i = 1, \dots, G$ and $j = 1, \dots, n_i$, were assumed to follow independent normal distributions with means of zero. In the second scenario, the random effects were assumed to have independent log-gamma distributions with zero means. In the two simulation scenarios, 15 groups were used with 2 or 5 subgroups. The sample sizes were 10, 40, or 100 individuals per subgroup. As for simulation studies I and II, two levels of variance components at both the subgroup and group levels were considered with the largest variance for the subgroup level as often found in the real data. Further, two amounts of censoring were used for each scenario, high and moderate censoring. Table 5.1 describes model parameter settings for the two scenarios of the simulations.

5.6.2. Simulation of data

A total of 1000 datasets were generated using R software in each setting of simulation study I, and 300 datasets in each setting of simulation studies II and III, because these simulations are computationally intensive and time consuming. Data for each particular setting were generated as follows: first, the random effects were generated from the considered probability density distribution. For the log-gamma distribution in simulation study I, the parameters ν and η were set, respectively, at 5.4834 and 4.9917 when the between-group variance was 0.2, and at 2.4599 and 1.9804 when $\sigma^2 = 0.5$. These values satisfied that $E(b_i) = \Psi(\nu) - \log\eta = 0$, and $var(b_i) = \Psi'(\nu) = 0.2$ and 0.5, respectively. Similarly, the parameters ν and η of a log-gamma distribution in simulation study III were set, respectively, at 1.4262 and 0.9657 when $\sigma_1^2 = 1$, and at 4.4793 and 3.9897 when σ_2^2 was 0.25. Furthermore, the scale parameter of Laplace distribution in simulation study I was set at 0.3162 and 0.5 to yield variances of 0.2 and 0.5, respectively. Next, the predictors were fixed across simulations and all binary with a 50%–50% division for variables at the first and third hierarchical levels, and with a 30%–70% division for the second level variable. In two-level hierarchical data, for example, the event time T_{ij} for each individual was randomly generated using the formula provided by Bender et al. (2005):

$$T_{ij} = \left[-\log(U)/\gamma \exp(\phi_{ij})\right]^{1/\alpha} \quad (5.7)$$

where U is a uniform variable on [0,1] and ϕ_{ij} is the linear predictor taking the form $b_i + \beta_1 x_{1ij} + \beta_2 x_{2i}$ in a model with random group effects. The censoring time C_{ij} was generated from a uniform distribution on [1,180]. The actual time for each individual was $Y_{ij} = T_{ij}$ with $\delta_{ij} = 1$ if $T_{ij} \leq C_{ij}$, or otherwise $Y_{ij} = C_{ij}$ with $\delta_{ij} = 0$.

5.6.3. Analysis and summary statistics of simulations

All estimation analysis for the simulation data was performed in Stata 12 using maximum likelihood estimation with Gaussian quadrature. The Cox model with random group effects was fit to simulated datasets using the xtpoisson command and a default number of integration points of 12, while the random coefficient and nested random effects Cox models were fit using xtmepoisson and a default of 7 integration points. Further, a fourth-order polynomial function of time was used for modeling the baseline hazard in all analyses. For each of the simulation iterations, model parameter estimates and their standard errors were extracted. Because of highly skewed distributions of the estimators from the simulations, particularly of variance components, we present simulation results as follows: the median for all model parameters, computed as the median of the estimated values across the simulated datasets; empirical and estimated standard errors (SE) for fixed effects, computed as the standard deviation and the mean of the model-based standard errors across the simulated datasets; the confidence interval (CI) probability coverage, computed as the proportion of simulations with the true value lies inside 95% CIs for fixed effect estimates in each analysis; the interquartile range (IQR) represented by the lower and upper quartiles for the distribution of variance component estimates; significance of a difference between median estimates and true values was tested using the Wilcoxon signed-rank and sign tests for fixed effects and variance component estimates, respectively; the z-test for proportions was used for the inference of CI coverage; significance of a difference between median estimates of correctly specified and misspecified models was assessed by a permutation test with 1,000 random permutations. When the estimation procedure failed to reach convergence in a particular analysis or dataset, the number of integration points of Gaussian quadrature method was changed (mostly increased), and the analysis was rerun until convergence was achieved. If non-sensible estimates (estimates that were >10 times or <10 times the true values) were

produced in a certain analysis, results were excluded from the statistics computed across the simulated datasets.

5.7. Simulation results

5.7.1. Simulation study I

The simulation results based on a Cox model with random group effects and two predictors, at individual and group levels, are shown in Tables 5.2, 5.3, 5.4 and 5.5.

5.7.1.1. Correctly specified distribution of random effects

In general, the fixed effect coefficient at the individual level, β_1, and its SE were estimated well when the censoring rate was 85% and the probability coverage was close to nominal. However, β_1 was estimated with mild downward bias when the censoring rate was 50% (absolute relative bias 1.6-3.4%); the estimated and empirical SEs agreed closely and the CI coverage was close to nominal with the exception of the largest data setting with $\sigma^2 = 0.5$. The fixed effect coefficient at group level, β_2, showed similar downward biases for a large number of groups with relatively large group sizes (100 and ≥ 40 for high and moderate censoring, respectively). When the number of groups (G) was large, the estimated and empirical SEs for β_2 estimates agreed closely and the CI coverage were close to nominal, whereas slightly underestimated SE and CI under-coverage were observed for $G = 20$. The between-group variance, σ^2, was estimated with a pronounced downwards bias in all settings. This bias decreased with increasing number of groups and group size.

When both the number of groups and the group size were small, the simulations reported a considerable amount of lower boundary estimates for σ^2, resulting in many zero SEs for these estimates. This problem was observed mostly in settings with a small magnitude of heterogeneity and a high censoring rate. Figure 5.3 displays the variance estimates on

logarithmic scale against their SEs for datasets with 85% censoring. For small data settings (top row of Figure 5.3), some estimates of $\log \sigma^2$ were less than -10, and most of these estimates had extreme SEs while the remaining estimates ($\log \sigma^2 > -5$) had SEs decreasing almost deterministically with increasing the $\log \sigma^2$ estimates. For large data dimensions (bottom row of Figure 5.3), the $\log \sigma^2$ estimates were larger compared with the smallest data settings and their SEs were clearly improved especially when $\sigma^2 = 0.5$. However, the functional relationship between $\log \sigma^2$ estimates and SEs remained. This was the reason for not providing CI probability coverage for variance components because the independence assumption of the normal and chi-square distributions in the numerator and denominator of the t-statistic needed for calculating CIs may be violated. Practically, when a zero group-level variance is encountered, the SE of such an estimate may be of less interest because the data analyst would deal with this situation as indicating that this was no detectable difference between groups.

5.7.1.2. Misspecified distribution of random effects

The results showed that the fixed effect estimates, β_1 and β_2, were estimated well regardless of the misspecification of random effects distribution, and in some cases the misspecified model produced estimates closer to the true value than those from the correct model. The agreement in β_1 estimates between the correct and misspecified models for a high censoring rate and a large magnitude of variance is shown in Figure 5.4. Table 5.6 presents the rejection rates (type I error and a test power) for testing different null hypothesis values for the parameters β_1 and β_2, in the simulation settings with $G = 70$ groups of $n_i = 100$. The results showed a close correspondence between the rejection rates of β_1 from the correct model and misspecified models and some slight differences in these rejection rates for β_2. Further, the lower censoring rate was associated with a higher power.

Generally speaking, the results showed that σ^2 estimates based on the log-gamma random effects model were larger and closer to the true value than those from the correct model in many cases (Figure 5.5). This was most obvious for $\sigma^2 = 0.5$ and 50% censoring (Table 5.5). In contrast, the opposite picture was observed for σ^2 estimates based on the Laplace random effects model which were mostly smaller and further away from the true value than those of the correctly specified model (Figure 5.5). Furthermore, there was almost no impact of misspecification on σ^2 estimates when censoring rate was 85% and $\sigma^2 = 0.2$. Additionally, the IQRs for σ^2 estimates agreed closely across all random effects distributions (Table 5.2). The IQRs for σ^2 estimates in other settings were quite different. For instance, the smallest width of IQRs was mostly for the correct model while the largest was for the Laplace random effects model with exception of the simulation settings of 50% censoring and $\sigma^2 = 0.5$ where the log-gamma random effects model showed the largest IQR widths (Table 5.5).

5.7.2. Simulation study II

The results were based on a Cox model with correlated random intercept and random slope. In each simulation setting, normal and Laplace distributions were assumed for random effects. The results of both the normal and Laplace random effects models are presented in Tables 5.7 and 5.8.

5.7.2.1. Correctly specified distribution of random effects

The fixed effect at individual level, β_1, was estimated well except for some cases. For instance, β_1 was estimated with upwards bias in the cases of $n_i = 10$ and a censoring rate of 85% (absolute relative bias 6.4-26.4%) and estimated with slightly downward bias for moderate censoring and a large magnitude of variances when $G = 70$ and $n_i \geq 40$ (absolute relative bias 3-6%). Overall the estimates of β_1 improved with increasing group size. The CI coverage of $\beta_1$was underestimated for $n_i = 100$ and overestimated for small group sizes. In

fact, an overestimation of the CI probability coverage was observed in some cases even when the estimated and empirical SEs were fairly close. No explanation of this finding can be offered except that the data maybe provided an insufficient amount of information as this phenomenon only appeared in the cases of $n_i \leq 40$ and $n_i = 10$ for 85% and 50% censoring, respectively. The group-level effect, β_2, was estimated reasonably well with best accuracy when $G = 70$. The SEs of β_2 estimates were underestimated for $G = 20$ and agreed closely with the empirical SEs when $G = 70$. Further, CI under-coverage was observed for β_2 estimates in the settings involving a small number of groups.

The random intercept variance, σ_0^2, was estimated with pronounced downwards bias. This bias was smaller for $G = 70$ than for $G = 20$. When the censoring rate was 85%, the random slope variance, σ_1^2, was estimated reasonably well for large magnitudes of heterogeneity and also for small magnitudes of heterogeneity but with $G = 70$. Otherwise, σ_1^2 was estimated with pronounced bias, mostly downwards. In general, the bias in σ_1^2 estimates increased with the magnitude of heterogeneity. For instance, the absolute relative bias of σ_1^2 estimates ranged between 1.2-16% and 3-30.2% for small and large magnitudes of heterogeneity, respectively. Finally, the correlation parameter, ρ, was estimated close to the true value in a few cases, in particular when $G = 70$ and $n_i \geq 40$, and it was estimated, otherwise, with strong biases, mostly upwards.

5.7.2.2. Misspecified distribution of random effects

The simulation results showed that misspecification of the random effects distribution had no impact on β_1 and β_2 estimates except for a few cases of β_1 estimation, but with no clear pattern. For instance, β_1 estimates under the misspecified model were clearly larger compared with those under the correct model for $G = 20$ and $n_i \geq 40$ when both the magnitude of heterogeneity and the censoring rate were large. In contrast, misspecifying the

random effects distribution had an impact on the estimation of variance components in a variety of situations. For example, the impact of misspecification on σ_0^2 estimates was observed mostly for 50% censoring rate and a large magnitude of heterogeneity, where σ_0^2 estimates were smaller (and away from the true value) for the misspecified model than the correct model. The impact of misspecification was also observed for σ_1^2 estimates. For instance, the estimates of σ_1^2 based on the misspecified model were smaller and away from the true value compared with σ_1^2 estimates of the correct model for a large magnitude of variance and $n_i \geq 40$ (Figure 5.6), or when both the magnitude of variance and the censoring rate were small. Finally, the correlation between the random intercept and random slope, ρ, was also influenced by misspecification of the random effects distribution in some cases. For example, the estimates of ρ were smaller under the misspecified model than for the correct model in the cases of 50% censoring and a large magnitude of variance, and also for small magnitude of variance but with $G = 70$ (Figure 5.6). The IQRs for σ_0^2 and ρ estimates under the misspecified model were larger compared with those corresponding to the correct model while the IQRs for σ_1^2 estimates from the misspecified and correct model were fairly similar.

5.7.3. Simulation study III

The simulation results based on a Cox model with nested random effects and three predictors, at different hierarchical levels, are shown in Table 5.9 and Table 5.10.

5.7.3.1. Correctly specified distribution of random effects

The lowest-level fixed effect parameter, β_1, was estimated well for the 85% censoring rate and also for the 50% censoring rate and a small magnitude of heterogeneity. Otherwise, it was estimated with downwards bias. The estimates of SE for β_1 were close to the observed SEs and their CI coverage was close to nominal except for some cases of $n_{ij} = 100$ where CI

under-coverage was observed. The fixed effects at the top levels, β_2 and β_3, were estimated close to the true values but their SEs were underestimated in almost all cases resulting in CI under-coverage (coverage ranged between 0.864-0.948 and 0.880-0.959 for β_2 and β_3, respectively).

The between-subgroup variance, σ_1^2, was estimated with pronounced biases in many cases. For instance, σ_1^2 was consistently underestimated for large magnitudes of heterogeneity and 50% censoring and also for small dimensions of data when the magnitude of heterogeneity was small. The between-group variance, σ_2^2, was estimated also with strong downwards biases (absolute relative biases 29.6-85.2% and 30.8-48.8% for heavy and moderate censoring, respectively) mostly due to small data dimensions and/or rare events as a high event rate provides more power to estimate variance components. These biases may be attributable to the small number of groups used in the study (15 groups) as an additional simulation study with 50 groups (2 subgroups per group), 85% censoring, and $\sigma_2^2 = 0.25$ showed a better performance. In that simulation, the median estimates for σ_2^2 based on $n_{ij} = 10$ and 100 were, respectively, 0.183 and 0.217. Finally, similar to simulation study I, estimates very close to zero for variance components, especially for σ_2^2, were obtained for small magnitudes of heterogeneity and 85% censoring, resulting in many zero SEs for these estimates (details not shown, but Tables 5.9 and 5.10 show $Q_1 = 0$ for many settings).

5.7.3.2. Misspecified distribution of random effects

With the exception of some very limited cases for β_2 and β_3 estimation, misspecification of the random effects distribution had no impact on the fixed effect estimates. However, the variance component estimates were clearly affected by misspecifying the random effects distribution. For instance, in the settings of 50% censoring, as well as in the cases of 85% that had relatively large subgroup sizes, larger σ_1^2 estimates for the misspecified model than for

the correctly specified model were observed, and the difference in the estimates between the two models increased with the magnitude of heterogeneity, event rate, and the dimension of data. Similarly, the estimates of σ_2^2 were larger under misspecified random effects distribution compared with the correct distribution of random effects. The IQRs for σ_1^2 and σ_2^2 estimates under the misspecified model were larger than the corresponding ranges of the correctly specified model in most of the settings.

5.8. Discussion

5.8.1. Estimation of fixed effect coefficients

The fixed effect coefficient at the individual level, β_1, was estimated well based on the three considered mixed-effects Cox models, regardless of the distribution of random effects in a wide variety of commonly encountered situations. However, the random slope and nested frailty models, particularly with correctly specified distribution of random effects, tended to produce estimates for β_1 with CI under-coverage when group/subgroup sizes were large. Similar findings based on a shared gamma frailty model and a Cox model with independent random intercept and random slope were reported in Glidden and Vittinghoff (2004) and Cortiñas et al. (2007), respectively. The fixed effect β_2 was also estimated quite well, especially when the number of groups was large, and it was not affected by misspecification of the random effects distribution. Furthermore, the estimation of β_3 under the nested frailty model was found to be robust, even though the results indicated that a modest bias can occur for a large magnitude of variance components and high dimensions of data. The estimated and empirical SEs of β_1 were similar under the true and misspecified random effects distribution and agreed closely in each scenario, whereas the SE of β_2 and β_3 estimates in some cases underestimated the empirical variability. Further, the CI coverage for β_2 and β_3 estimates was mostly under nominal, and the rejection rates of fixed effects based on Cox

models with normal random group effects were close to those from the same models with truly log-gamma and Laplace distributions for random effects (Neuhaus et al., 2011).

5.8.2. Estimation of random effect variances

The overall impression was that the Poisson approach for mixed-effects Cox models tended to underestimate the true underlying heterogeneity. Similar results were reported in Crowther et al. (2012). In the Cox model with random group effects, misspecifying the distribution of random effects resulted in a non-robust estimation of the between-group variance σ^2, especially for moderate censoring rate or a large magnitude of variance components. When the random effects followed a log-gamma distribution, the estimates of σ^2 were found to be larger and in many cases closer to the true values than the estimates of the normal random effects model. In contrast, the estimates based on the Laplace random effects model were almost always smaller and further apart from the true values compared with those based on the normal random effects model. Using a gamma frailty model, Duchateau and Janssen (2008) pointed out that increasing the frailty variance led to a large bias in the heterogeneity parameter when the frailties were log-normally distributed. Similar findings were previously reported by Ferreira and Gracia (2001) and Massonnet et al. (2006).

For the variance components estimation based on the Cox model with correlated random intercept and random slope, the bias in the estimates of variance components was clearly elevated under the Laplace random effects model compared with the correctly specified model, especially in the random slope variance σ_1^2 and the correlation parameter ρ for moderate censoring rate, and a heavy censoring with large group sizes. Generally speaking, the estimates of variance components based on the Laplace random effects model were smaller and further apart from the true values than was the case for the normal random effects

model. Cortiñas et al. (2007) reported similar findings based on the same true and analysis models but with independent random effects.

The estimation of the variance components based on the Cox model with nested random effects was also found to be non-robust and misspecification of the random effects distribution had a larger impact on the between-subgroup variance σ_1^2 than on the between-group variance σ_2^2 estimates. There was a considerable amount of upwards bias in the estimates of σ_1^2 when the true random effects distribution was log-gamma compared with normal random effects, and this bias increased with the magnitude of variance, and dimension of data. Furthermore, the point estimates of the heterogeneity parameter σ_2^2 based on the log-gamma random effects model were somewhat larger than those based on the normal random effects model for large magnitudes of heterogeneity, event rate, and data dimensions, although σ_2^2 was strongly underestimated in the two scenarios. This underestimation of σ_2^2 was possibly due to the small number of units at the top level as an extra simulation study with a larger number of top-level units showed a clear improvement.

For small magnitudes of heterogeneity and heavy censoring, the IQRs for the variance component estimates produced by the misspecified and correct models were comparable under the shared and nested frailty models. Otherwise, the IQRs of these estimates were larger in the misspecified model than in the correct model. Taking into account the above, misspecification of the random effects distribution might be an issue when fitting mixed effects Cox models because invalid estimates for variance components may occur.

Finally, it is important to mention that with rare events and a small number of groups with relatively small sizes, the estimation approach tended to estimate the variance components and their SEs as zero. The degree to which such unrealistic estimates and SEs in the aforementioned settings occurred was substantial for the top-level variance in the shared and

nested frailty models. In practice, zero group-level variance and SE is often dealt with as indicating no difference between groups, although in some cases this may cause problems as recently shown in Chung et al. (2013). To avoid zero variance estimates, these authors suggest using maximum penalized likelihood with penalty from the log-gamma family for hierarchical linear models; this can also be applied to GLMMs.

5.8.3. Effect of the magnitude of heterogeneity

The magnitude of bias in the variance component estimates due to misspecification of the random effects distribution was larger for a large magnitude of heterogeneity than for a small magnitude of heterogeneity. This may be explained by smaller difference in shape between the assumed distribution (normal) and the true distribution (log-gamma or Laplace) for a small magnitude of variance than for a large variance. Therefore, the findings were in agreement with results in Duchateau and Janssen (2008) based on a shared frailty model and with findings of Sastry (1997) based on a nested frailty model.

5.8.4. Effect of censoring

The difference in the estimates of variance components between the correct and misspecified models was almost always larger for moderate censoring than for heavy censoring. In other words, the impact of misspecifying the distribution of random effects on variance estimates was strongest for high (50%) event rates. This applied to all considered models and one may hypothesize that increasing the event rate would provide more power to estimate complex random component structures, especially when both the number of groups and the group size are large enough.

5.8.5. Effect of data dimensions

In general, the number of groups/subgroups and group/subgroup sizes had an impact on model parameter estimation under both the misspecified and correctly specified random-effects distribution. The estimation of model parameters was mostly improved when the number and the size of groups/subgroups increased. When the random effects followed a log-gamma distribution, the absolute bias in the variance component estimates due to misspecifying the random effects distribution increased with the dimensions of data in the nested frailty model, and in the shared frailty model for a large magnitude of heterogeneity. In some cases of the shared frailty model, the between-group variance was estimated closer to the true values when fitting a misspecified model with log-gamma random effects than when fitting the correct model. For Laplace random effects, the absolute bias in σ^2 estimates under the shared frailty model increased with decreasing number and size of groups. In the random slope model, there was no a clear pattern for the magnitude of bias in variance estimates due to misspecification of the random effects density. This would make it hard to predict the direction of bias in variance component estimates in this model.

5.8.6. Model design

The effect of misspecification of the random effects distribution on the variance component estimation was minor among model designs and reflected by the number of parameters. For instance, under the same circumstances misspecifying the random effects distribution in the random slope model had a slightly larger impact on the random intercept variance estimates than the shared frailty model for a large number of groups with small sizes and nearly the same impact otherwise. Further, almost the same impact of misspecified the distribution of random effects was detected in the between-group and between-subgroup variances for the shared and nested frailty models, respectively, under the same frailty distribution, magnitude of heterogeneity, and censoring rate.

5.9. Concluding remarks

Via simulations, we have shown that the Poisson GLMM approach for estimating a mixed-effects Cox model, with normally distributed random effects, provides robust estimation for fixed effects under many circumstances even with misspecified distribution of random effects. Based on the results of current study, misspecification of random effects distribution has negligible impact on regression coefficient estimates. Practically, the estimation of fixed effects is often viewed as more important than the estimation of variance components, especially in survival analysis where the main purpose of accounting for data structure is often improving model estimation. On the other hand, misspecifying the random effects distribution may result in biased estimation for the variance components. The magnitude of bias in variance components due to misspecification under one-component random effects Cox models is somewhat comparable to the corresponding bias based on multi-component random effects Cox models when they are compared for the same levels. Finally, misspecification of the random effects distribution may have substantial impacts in situation having a large magnitude of variability between random effects, event rate, and data dimensions; in such cases caution should be exercised.

References

Aalen O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, **7**: 1121-1137.

Aalen O. (1992). Modelling heterogeneity in survival analysis by compound Poisson distribution. *Annals of Applied Probability*, **2**: 951-972.

Bender R, Augustin T, Blettner M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, **24**: 1712-1723.

Chung Y, Rabe-Hesketh S, Dorie V, Gelman A, Liu J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychomerika*, DOI:10.1007/S11336-013-9328-2.

Cortiñas J, Burzykowski T. (2005). A version of the EM algorithm for proportional hazard model with random effects. *Biometrical Journal*, **47**: 847-862.

Cortiñas J, Legrand C, Burzykowski T, Janssen P, Ducrocq V, Duchateau L. (2007). Comparison of different estimation procedures for proportional hazards model with random effects. *Computational Statistics & Data Analysis*, **51**: 3913-3930.

Crowther M, Riley R, Staessen J, Wang J, Gueyffier F, Lambert P. (2012). Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Medical Research Methodology*, **12**: 34. DOI: 10.1186/1471-2288-12-34. URL.

Duchateau L, Janssen P. (2008). *The Frailty Model*. New York: Springer.

Feng S, Wolfe R, Port F. (2005). Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *Journal of the American Statistical Association*, **100**: 718-735.

Feng S, Nie L, Wolfe R. (2009). Laplace's approximation for relative risk frailty models. *Lifetime Data Analysis*, **15**: 343-356.

Ferreira A, Garcia N. (2001). Simulation study for misspecifications on a frailty model. *Brazilian Journal of Probability and Statistics*, **15**: 121-134.

Glidden D, Vittinghoff E. (2004). Modeling clustered survival data from multicentre clinical trials. *Statistics in Medicine*, **23**: 369-388.

Hangagal D. (2011). *Modelling Survival Data Using Frailty Models*. Chapman & Hall/CRC.

Hougaard P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, **1**: 255-273.

Legrand C, Ducrocq V, Janssen P, Sylvester R, Duchateau L. (2005). A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model. *Statistics in Medicine*, **24**: 3789-3804.

Ma R, Krewski D, Burnett R. (2003). Random effects Cox models: A Poisson modelling approach. *Biometrika*, **90**: 157-169.

Massonnet G, Burzykowski T, Janssen. (2006). Resampling plans for frailty models. *Communications in Statistics- Simulation and Computation*, **35**: 497-514.

McGilchrist C, Aisbett C. (1991). Regression with frailty in survival analysis. *Biometrics*, **47**: 461-466.

Neuhaus J, McCulloch C, Boylan R. (2011). A note on type II error under random effects misspecification in generalized linear mixed models. *Biometrics*, **67**: 654-660.

Nielsen G, Gill R, Andersen P, Sorensen T. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, **19**: 25-43.

Rabe-Hesketh S, Skrondal A. (2012). *Multilevel and Longitudinal Modeling using Stata*. *Volume II: Categorical Responses, Counts, and Survival*, 3rd ed, Stata Press.

Ripatti S, Palmgren J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56**: 1016-1022.

Rondeau V, Filleul L, Joly P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine,* **25**: 4036-4052.

Rondeau V, Michiels S, Liquet B, Pignon J. (2008). Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine,* **27**: 1894-1910.

Sastry N. (1997). A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of the American Statistical Association*, **92**: 426-435.

Stryhn H, Christensen J. (2014). The analysis – hierarchical models: past, present and future. *Preventive Veterinary Medicine,* **113**: 304-312.

Therneau T, Grambsch P. (2000). *Modelling Survival Data, Extending the Cox model*. New York: Springer.

Waldner C. (2008). Western Canada study of animal health effects associated with exposure to emissions from oil and natural gas field facilities. Study design and data collection I. Herd performance records and management. *Archives of Environmental & Occupational Health No 4*. **63**: 167-184.

Wienke A. (2010). *Frailty Models in Survival Analysis.* Boca Raton: Taylor & Francis.

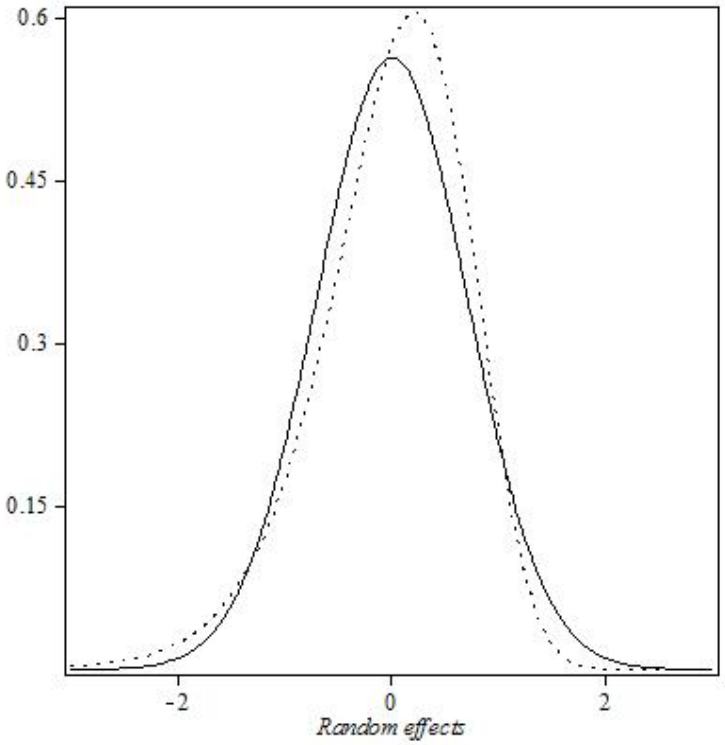

Figure 5.1: The normal (solid) and log-gamma (dot) probability densities with mean zero and variance σ^2 = 0.5.

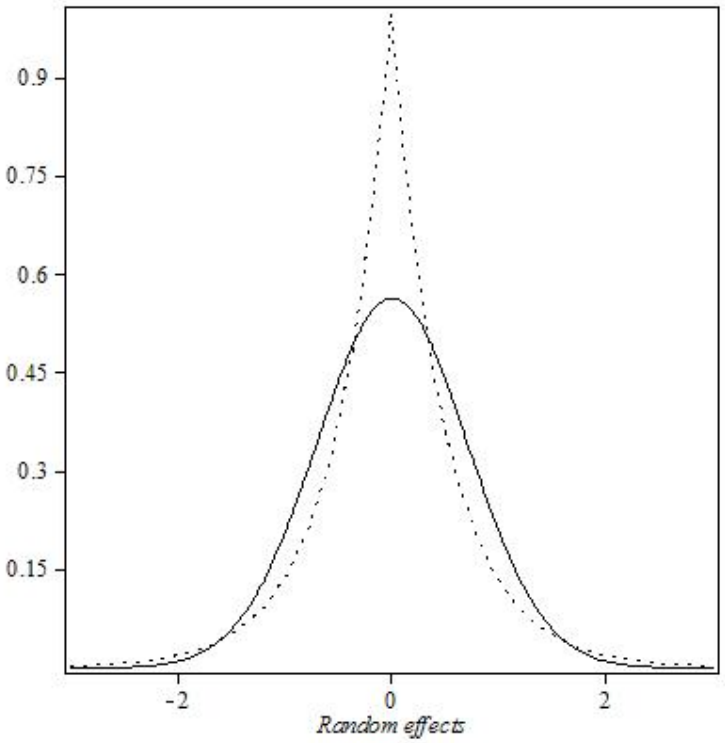

Figure 5.2: The normal (solid) and Laplace (dot) probability densities with mean zero and variance $\sigma^2 = 0.5$.

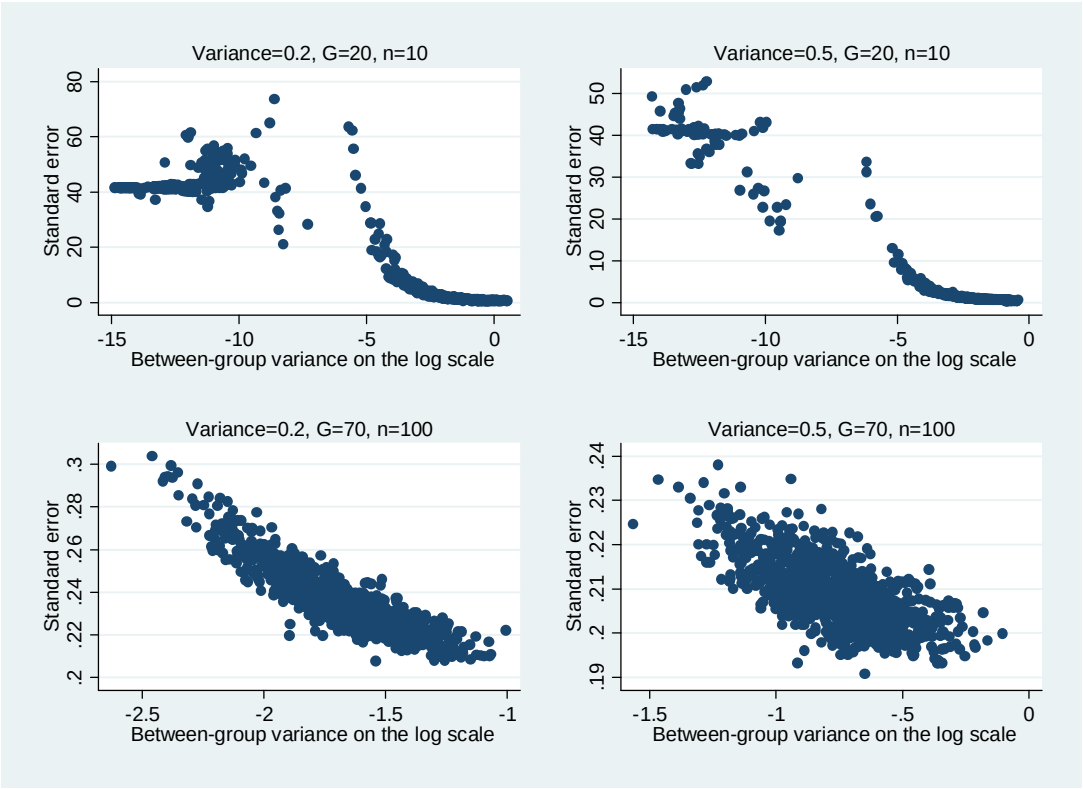

Figure 5.3. Scatter plots for ML estimates of between-group variance on the log scale vs. their standard errors, in a shared frailty model for settings of 85% censoring and different magnitudes of variance, number of groups, and group size.

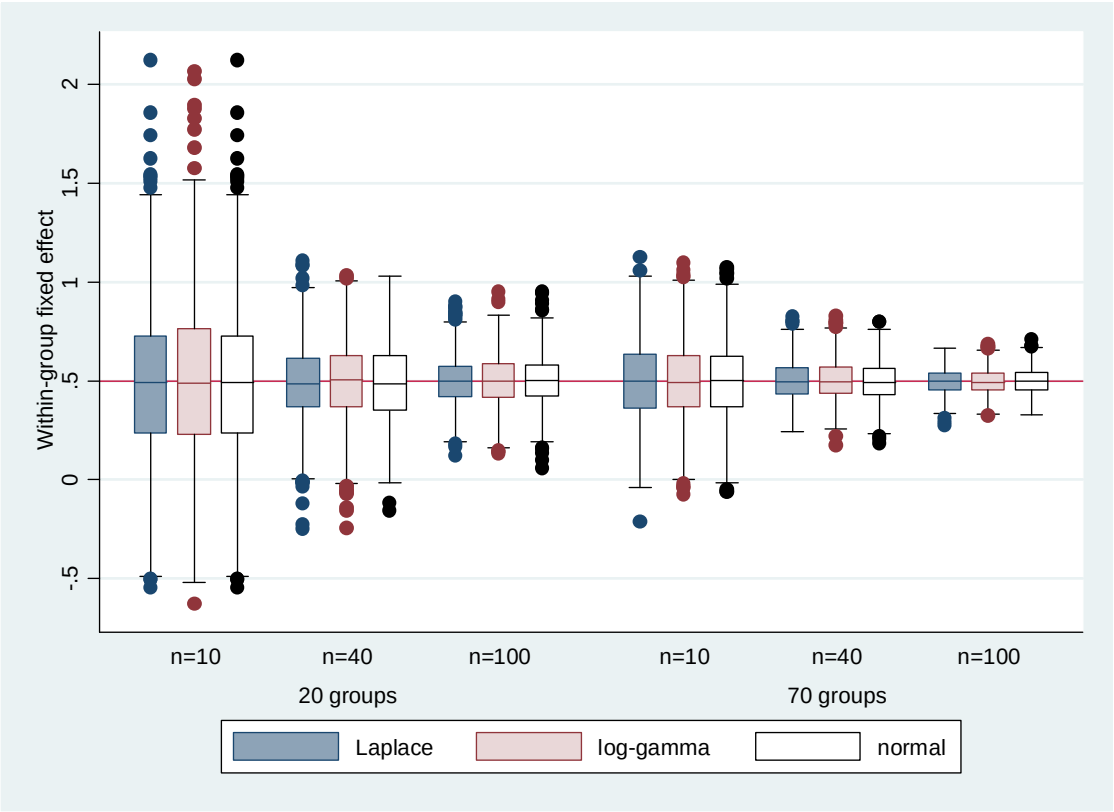

Figure 5.4. Boxplots of estimates of within-group fixed effect (β_1) from a shared frailty model with Laplace, log-gamma, and normal distributed random effects, fit with assumed normal distribution for random effects to datasets with the settings of 85% censoring; 20 and 70 groups, each with sizes of 10, 40, and 100; and $\sigma^2 = 0.5$. The reference line is at the true value of $\beta_1 = 0.5$.

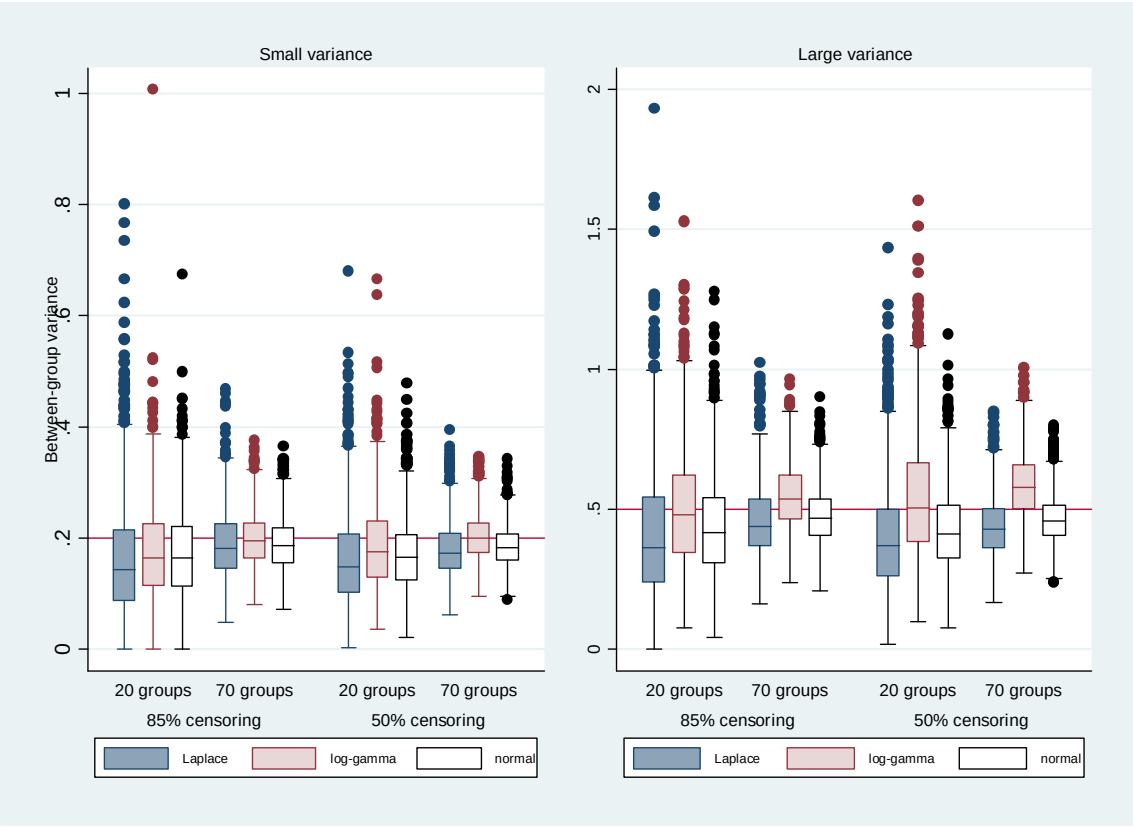

Figure 5.5. Boxplots of estimates of between-group variance (σ^2) from a shared frailty model with Laplace, log-gamma, and normal distributed random effects, fit with assumed normal distribution for random effects to datasets with the settings of 85% and 50% censoring; 20 and 70 groups with size of 100; small (0.2) and large (0.5) variance. The reference lines are at the true values of $\sigma^2 = 0.2$ and 0.5 for a small and large variance, respectively.

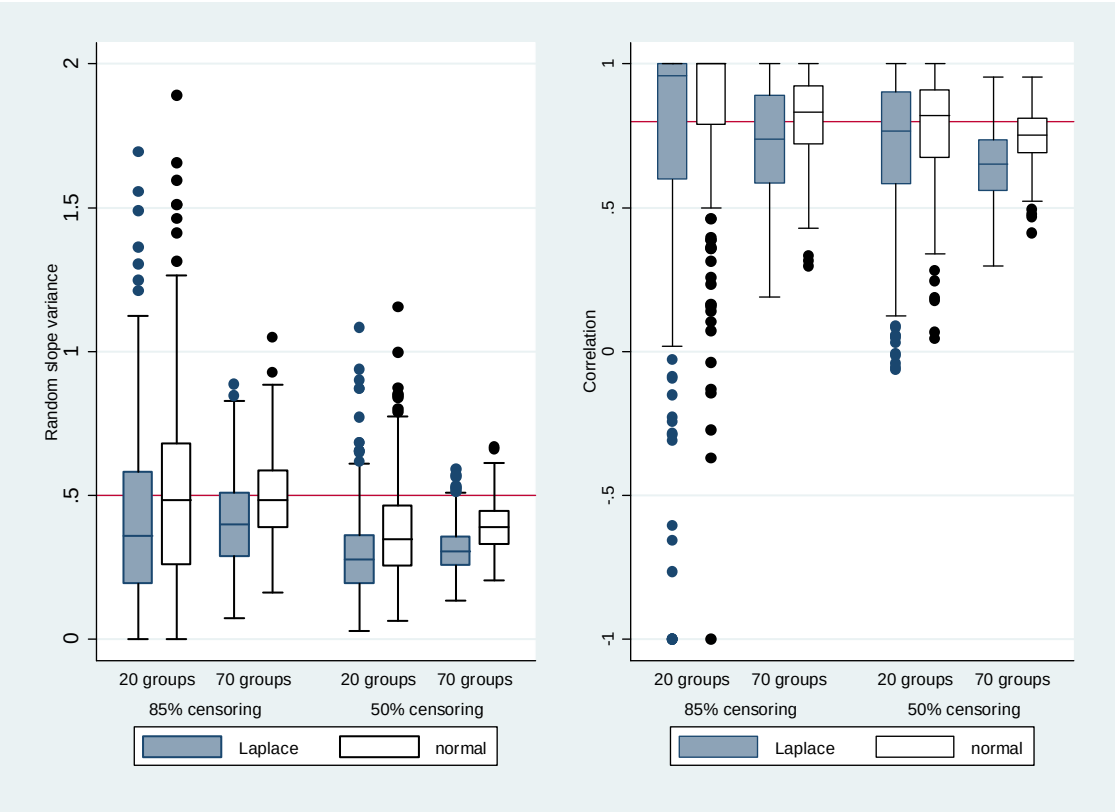

Figure 5.6. Boxplots of the estimates of random-slope variance (left) and correlation between the random intercept and random slope (right) from a random coefficient Cox model with Laplace and normal distributed random effects, fit with assumed normal distribution for random effects to datasets with the settings of 85% and 50% censoring rates; 20 and 70 groups with size of 100; and large variance components. The reference lines are at the true values of $\sigma_1^2 = 0.5$ and $\rho = 0.8$.

Table 5.1. True distributions and model parameter settings used for three simulation studies based on Cox model with random group effect (study I), Cox model with correlated random intercept and random slope (study II), and Cox model two nested random effects (study III).

Random effects dist'n	**85% censoring**				**50% censoring**				**Fixed effect parameters[a]**
	Small variance		**Large variance**		**Small variance**		**Large variance**		
	(α, γ)[c]	**var. parms[b]**	**(α, γ)[c]**	**var. parms[b]**	**(α, γ)[c]**	**var. parms[b]**	**(α, γ)[c]**	**var. parms[b]**	
Simulation study I									
Normal	(0.5, 0.012)	(0.2)	(0.5, 0.010)	(0.5)	(0.5, 0.060)	(0.2)	(0.5, 0.050)	(0.5)	(0.5, 0.3)
Log-gamma	(0.5, 0.015)	(0.2)	(0.5, 0.016)	(0.5)	(0.5, 0.070)	(0.2)	(0.5, 0.090)	(0.5)	
Laplace	(0.5, 0.012)	(0.2)	(0.5, 0.010)	(0.5)	(0.5, 0.060)	(0.2)	(0.5, 0.050)	(0.5)	
Simulation study II									
Normal	(0.5, 0.008)	(0.5, 0.25, 0.4)	(0.5, 0.006)	(1.0, 0.50, 0.8)	(0.5, 0.050)	(0.5, 0.25, 0.4)	(0.5, 0.050)	(1.0, 0.50, 0.8)	(0.5, 0.3)
Laplace	(0.5, 0.008)	(0.5, 0.25, 0.4)	(0.5, 0.006)	(1.0, 0.50, 0.8)	(0.5, 0.050)	(0.5, 0.25, 0.4)	(0.5, 0.050)	(1.0, 0.50, 0.8)	
Simulation study III									
Normal	(0.5, 0.008)	(0.5, 0.25)	(0.5, 0.006)	(1.0, 0.5)	(0.5, 0.044)	(0.5, 0.25)	(0.5, 0.042)	(1.0, 0.5)	(0.5, 0.3, 0.2)
Log-gamma	(0.5, 0.015)	(0.5, 0.25)	(0.5, 0.023)	(1.0, 0.5)	(0.5, 0.095)	(0.5, 0.25)	(0.5, 0.160)	(1.0, 0.5)	

[a] Fixed effect parameters in each simulation study: I and II: (β_1, β_2); III: $(\beta_1, \beta_2, \beta_3)$.

[b] Variance parameter(s) in each simulation study: I: (σ^2); II: $(\sigma_0^2, \sigma_1^2, \rho)$; III: (σ_1^2, σ_2^2).

[c] Baseline hazard parameters.

Table 5.2. Poisson modeling results for data generated from Cox models with **85% censoring rate** and random group effects with $\boldsymbol{\sigma^2 = 0.2}$ and 3 different assumed random effects distributions: median of point estimates, empirical standard error and mean of the estimated standard error for fixed effects as well as probability coverage, and interquartiles ranges [$\mathbf{Q_1}$, $\mathbf{Q_3}$] for variance component estimates, over 1000 simulated datasets.

True r.e.			β_1		β_2		σ^2
dist'n	G	n_i	**median (std; se)**	**Pr. C.**	**median (std; se)**	**Pr. C.**	**median [Q_1; Q_3]**
True	**20**		**.500**		**.300**		**.200**
Normal		**10**	.478 (.372; .377)	$.966^{+}$	.306 (.444; .428)	.943	$.063^{-}$ [.000; .240]
Log-gamma			$.525^{+}$ (.382; .375)*	.963	.303 (.461; .426)	.944	$.056^{-}$ [.000; .246]
Laplace			.494 (.409; .380)	.937	.283 (.447; .437)	.950	$.074^{-}$ [.000; .265]
Normal		**40**	.490 (.194; .187)	.941	$.269^{-}$ (.303; .268)	$.918^{-}$	$.144^{-}$ [.072; .228]
Log-gamma			.489 (.180; .184)	.948	.304 (.291; .268)*	$.911^{-}$	$.148^{-}$ [.081; .234]
Laplace			.487 (.185; .186)	.948	.291 (.287; .265)	$.919^{-}$	$.135^{-}$ [.057; .227]
Normal		**100**	.497 (.117; .116)	.948	.305 (.245; .227)	$.920^{-}$	$.164^{-}$ [.112; .221]
Log-gamma			.504 (.113; .115)	.946	.315 (.249; .228)	$.912^{-}$	$.164^{-}$ [.115; .226]
Laplace			.503 (.118; .117)	.948	.299 (.245; .221)	$.912^{-}$	$.143^{-}$ [.088; .215]*
	70						
Normal		**10**	.502 (.182; .190)	.963	.289 (.238; .234)	.941	$.155^{-}$ [.065; .256]
Log-gamma			.499 (.189; .188)	.948	.297 (.229; .232)	.957	$.168^{-}$ [.071; .257]
Laplace			.500 (.188; .191)	.960	.307 (.240; .238)	.955	$.177^{-}$ [.082; .282]*
Normal		**40**	$.492^{-}$ (.095; .096)	.956	.293 (.152; .148)	$.936^{-}$	$.182^{-}$ [.145; .219]
Log-gamma			.507 (.093; .095)*	.956	.294 (.145; .148)	.954	$.182^{-}$ [.144; .223]
Laplace			.496 (.094; .097)	.960	.295 (.158; .150)	.941	$.185^{-}$ [.139; .241]
Normal		**100**	.499 (.060; .060)	.953	$.285^{-}$ (.124; .127)	.947	$.187^{-}$ [.156; .219]
Log-gamma			.501 (.059; .059)	.962	.298 (.125; .129)	.948	$.195^{-}$ [.163; .228]*
Laplace			.499 (.059; .060)	.950	.295 (.133; .127)	.940	$.182^{-}$ [.146; .226]

* Significantly different from the correctly specified model.
$^{-}$ Significantly underestimated and $^{+}$ overestimated from the true value.

Table 5.3. Poisson modeling results for data generated from Cox models with **50% censoring rate** and random group effects with $\boldsymbol{\sigma^2 = 0.2}$ and 3 different assumed random effects distributions: median of point estimates, empirical standard error and mean of the estimated standard error for fixed effects as well as probability coverage, and interquartiles ranges [$\mathbf{Q_1}$, $\mathbf{Q_3}$] for variance component estimates, over 1000 simulated datasets.

True r.e. dist'n	G	n_i	β_1 **median (std; se)**	**Pr. C.**	β_2 **median (std; se)**	**Pr. C.**	σ^2 **median [Q_1; Q_3]**
True	**20**		**.500**		**.300**		**.200**
Normal		**10**	.497 (.201; .203)	.950	.290 (.284; .275)	.944	.139$^-$ [.064; .228]
Log-gamma			.488 (.205; .206)	.949	.318 (.298; .277)	.922$^-$	.139$^-$ [.065; .224]
Laplace			.495 (.205; .203)	.939	.289 (.285; .273)	.930$^-$	.130$^-$ [.049; .217]
Normal		**40**	.492$^-$ (.100; .101)	.957	.284$^-$ (.240; .215)	.916$^-$	.156$^-$ [.114; .207]
Log-gamma			.493 (.100; .102)	.954	.300 (.238; .223)	.930$^-$	.168$^-$ [.123; .227]*
Laplace			.493$^-$ (.098; .101)	.960	.294 (.230; .208)	.924$^-$	.140$^-$ [.094; .201]*
Normal		**100**	.490$^-$ (.064; .063)	.946	.299 (.217; .202)	.954	.165$^-$ [.125; .207]
Log-gamma			.493$^-$ (.063; .064)	.950	.305 (.227; .210)	.922$^-$	.175$^-$ [.130; .231]*
Laplace			.490$^-$ (.061; .063)	.957	.290 (.212; .196)	.923$^-$	.149$^-$ [.101; .207]*
	70						
Normal		**10**	.488$^-$ (.102; .105)	.945	.293 (.153; .153)	.946	.168$^-$ [.130; .216]
Log-gamma			.487$^-$ (.103; .107)	.958	.290 (.154; .156)	.949	.179$^-$ [.134; .223]*
Laplace			.482$^-$ (.100; .106)	.954	.296 (.157; .151)	.935	.155$^-$ [.118; .206]*
Normal		**40**	.488$^-$ (.051; .053)	.948	.291$^-$ (.122; .120)	.936	.179$^-$ [.154; .208]
Log-gamma			.492$^-$ (.052; .053)	.952	.287$^-$ (.124; .124)	.946	.192$^-$ [.167; .223]*
Laplace			.493$^-$ (.050; .053)*	.947	.287 (.127; .118)	.935$^-$	.171$^-$ [.140; .206]*
Normal		**100**	.489$^-$ (.032; .033)	.939	.284$^-$ (.111; .113)	.948	.183$^-$ [.161; .208]
Log-gamma			.491$^-$ (.033; .033)	.943	.295 (.118; .118)	.942	.200 [.173; .228]*
Laplace			.488$^-$ (.032; .033)	.950	.290$^-$ (.113; .111)	.941	.174$^-$ [.147; .208]*

* Significantly different from the correctly specified model.
$^-$ Significantly underestimated and $^+$ overestimated from the true value.

Table 5.4. Poisson modeling results for data generated from Cox models with **85% censoring rate** and random group effects with $\boldsymbol{\sigma^2 = 0.5}$ and 3 different assumed random effects distributions: median of point estimates, empirical standard error and mean of the estimated standard error for fixed effects as well as probability coverage, and interquartiles ranges [$\mathbf{Q_1}$, $\mathbf{Q_3}$] for variance component estimates, over 1000 simulated datasets.

True r.e. dist'n	G	n_i	β_1 **median (std; se)**	**Pr. C.**	β_2 **median (std; se)**	**Pr. C.**	σ^2 **median [Q_1; Q_3]**
True	**20**		**.500**		**.300**		**.500**
Normal		**10**	.493 (.379; .391)	.963	.300 (.548; .503)	.931^{-}	.278^{-} [.061; .559]
Log-gamma			.488 (.398; .394)	.960	.304 (.541; .507)	.942	.290^{-} [.089; .577]
Laplace			.483 (.419; .397)	.946	.283 (.553; .524)	.943	.317^{-} [.071; .652]*
Normal		**40**	.484^{-} (.200; .194)	.949	.272^{-} (.419; .370)	.909^{-}	.396^{-} [.252; .567]
Log-gamma			.505 (.198; .192)	.942	.285 (.412; .374)	.915^{-}	.402^{-} [.263; .575]
Laplace			.485 (.194; .194)	.947	.305 (.394; .362)	.924^{-}	.358^{-} [.213; .545]*
Normal		**100**	.502 (.121; .121)	.956	.309 (.361; .336)	.929^{-}	.417^{-} [.308; .541]
Log-gamma			.497 (.122; .119)	.946	.313 (.380; .355)	.919^{-}	.480^{-} [.344; .622]*
Laplace			.498 (.121; .121)	.947	.302 (.357; .323)	.912^{-}	.363^{-} [.241; .545]*
	70						
Normal		**10**	.501 (.191; .197)	.959	.290 (.289; .281)	.946	.439^{-} [.299; .595]
Log-gamma			.492 (.197; .197)	.943	.306 (.285; .276)	.942	.412^{-} [.296; .532]*
Laplace			.497 (.197; .198)	.957	.320 (.300; .290)*	.947	.495 [.335; .674]*
Normal		**40**	.493 (.101; .100)	.947	.289 (.210; .207)	.942	.461^{-} [.391; .543]
Log-gamma			.496 (.101; .099)	.948	.298 (.211; .210)	.950	.491 [.415; .586]*
Laplace			.495 (.096; .100)	.960	.298 (.217; .206)	.944	.463^{-} [.363; .571]
Normal		**100**	.498 (.062; .062)	.943	.282^{-} (.180; .188)	.949	.469^{-} [.405; .538]
Log-gamma			.492^{-} (.059; .061)	.965^{+}	.306 (.199; .199)*	.950	.538^{+} [.464; .622]*
Laplace			.497 (.063; .062)	.948	.296 (.192; .184)	.940	.440^{-} [.371; .536]*

* Significantly different from the correctly specified model.
$^{-}$ Significantly underestimated and $^{+}$ overestimated from the true value.

Table 5.5. Poisson modeling results for data generated from Cox models with **50% censoring rate** and random group effects with $\boldsymbol{\sigma^2 = 0.5}$ and 3 different assumed random effects distributions: median of point estimates, empirical standard error and mean of the estimated standard error for fixed effects as well as probability coverage, and interquartiles ranges [$\mathbf{Q_1}$, $\mathbf{Q_3}$] for variance component estimates, over 1000 simulated datasets.

True r.e. dist'n	G	n_i	β_1 **median (std; se)**	**Pr. C.**	β_2 **median (std; se)**	**Pr. C.**	σ^2 **median [Q_1; Q_3]**
True	**20**		**.500**		**.300**		**.500**
Normal		**10**	.498 (.211; .214)	.957	.296 (.391; .371)	.920$^-$	.375$^-$ [.239; .550]
Log-gamma			.483$^-$ (.205; .208)	.963	.306 (.401; .379)	.924$^-$	.406$^-$ [.274; .587]*
Laplace			.493 (.215; .215)	.957	.286 (.380; .360)	.932$^-$	.347$^-$ [.206; .516]*
Normal		**40**	.487$^-$ (.099; .101)	.950	.276$^-$ (.357; .320)	.920$^-$	.394$^-$ [.309; .516]
Log-gamma			.490$^-$ (.101; .103)	.954	.284$^-$ (.381; .348)	.922$^-$	.481$^-$ [.354; .632]*
Laplace			.486$^-$ (.102; .106)	.962	.291 (.336; .306)	.924$^-$	.351$^-$ [.249; .501]*
Normal		**100**	.488$^-$ (.067; .060)	.946	.314 (.338; .312)	.916$^-$	.413$^-$ [.323; .515]
Log-gamma			.486$^-$ (.063; .064)	.943	.317 (.369; .346)	.922$^-$	.506 [.384; .667]*
Laplace			.487$^-$ (.064; .066)	.956	.290 (.324; .298)	.926$^-$	.370$^-$ [.261; .501]*
	70						
Normal		**10**	.489$^-$ (.108; .111)	.955	.289 (.205; .207)	.946	.439$^-$ [.362; .524]
Log-gamma			.480$^-$ (.105; .108)	.948	.293 (.214; .212)	.944	.494 [.412; .582]*
Laplace			.480$^-$ (.105; .111)	.962	.297 (.209; .200)	.940	.399$^-$ [.322; .484]*
Normal		**40**	.483$^-$ (.054; .055)	.947	.282$^-$ (.182; .179)	.940	.452$^-$ [.393; .513]
Log-gamma			.483$^-$ (.052; .053)	.944	.286$^-$ (.199; .196)	.940	.553$^+$ [.476; .637]*
Laplace			.488$^-$ (.053; .055)	.946	.281$^-$ (.184; .173)	.932$^-$	.414$^-$ [.349; .492]*
Normal		**100**	.486$^-$ (.033; .034)	.931$^-$	.278$^-$ (.171; .174)	.938	.460$^-$ [.406; .515]
Log-gamma			.485$^-$ (.033; .033)	.932$^-$	.292 (.196; .194)	.942	.578$^+$ [.504; .659]*
Laplace			.485$^-$ (.033; .034)	.932$^-$	.282$^-$ (.172; .169)	.939	.428$^-$ [.362; .504]*

* Significantly different from the correctly specified model.
$^-$ Significantly underestimated and $^+$ overestimated from the true value.

Table 5.6. The rejection rates (type I and a test power) for testing different null hypothesis values for the parameters β_1 and β_2, in the simulation settings with $G = 70$ groups of $n_i = 100$ based on Cox models with random group effects and 3 different assumed random effects distributions.

	β_1				β_2			
H_0 values	**0.5**	**0.45**	**0.4**	**0.35**	**0.3**	**0.22**	**0.14**	**0.06**
$\sigma^2 = 0.2$, 85% censoring								
Normal	.047	.129	.396	.695	.053	.087	.207	.424
Log-gamma	.038	.151	.394	.714	.052	.090	.245	.456
Laplace	.050	.129	.376	.715	.060	.105	.237	.459
$\sigma^2 = 0.5$, 85% censoring								
Normal	.057	.120	.363	.673	.051	.070	.123	.214
Log-gamma	.035	.102	.352	.659	.050	.078	.138	.238
Laplace	.052	.123	.337	.676	.060	.078	.141	.262
$\sigma^2 = 0.2$, 50% censoring								
Normal	.061	.217	.787	.995	.052	.095	.237	.511
Log-gamma	.057	.216	.794	.989	.058	.098	.280	.508
Laplace	.050	.206	.775	.989	.059	.110	.274	.557
$\sigma^2 = 0.5$, 50% censoring								
Normal	.069	.180	.732	.988	.062	.071	.133	.233
Log-gamma	.068	.180	.725	.987	.058	.073	.143	.239
Laplace	.068	.153	.690	.977	.061	.075	.139	.280

Table 5.7. Poisson modeling results for data generated from random coefficient Cox models with **85% censoring rate** and two different assumed random effects distributions: median of point estimates, empirical standard error and mean of the estimated standard error for fixed effects as well as probability coverage, and interquartiles ranges [$\mathbf{Q_1}$, $\mathbf{Q_3}$] for variance component estimates, over 300 simulated datasets.

True r.e.			β_1		β_2		σ_0^2	σ_1^2	ρ
dist'n	G	n_i	**median (std; se)**	**Pr. C.**	**median (std; se)**	**Pr. C.**	**median [Q_1; Q_3]**	**median [Q_1; Q_3]**	**median [Q_1; Q_3]**
True	**20**		**.500**		**.300**		**.500**	**.250**	**.400**
Normal		**10**	$.581^{+}$ (.590; .621)	$.986^{+}$	$.338^{+}$ (.660; .637)	.949	$.415^{-}$ [.155; .907]	$.409^{+}$ [.064; 1.025]	$-.462^{-}$ [-1.00; 1.00]
Laplace			$.613^{+}$ (.580; .614)	$.987^{+}$	$.435^{+}$ (.732; .666)	$.933^{-}$	$.377^{-}$ [.104; .918]*	$.334^{+}$ [.041; .977]	$-.414^{-}$ [-1.00; 1.00]
Normal		**40**	.483 (.291; .288)	$.967^{+}$	.246 (.483; .455)	.946	$.383^{-}$ [.214; .647]	.246 [.099; .432]	$.921^{+}$ [-.124; 1.00]
Laplace			$.539^{+}$ (.279; .281)*	.946	.308 (.504; .453)	$.920^{-}$	$.407^{-}$ [.199; .666]	$.198^{-}$ [.067; .411]*	1.00^{+} [-.209; 1.00]
Normal		**100**	.498 (.207; .196)	$.936^{-}$	.277 (.418; .404)	$.930^{-}$	$.416^{-}$ [.282; .596]	$.213^{-}$ [.109; .368]	$.716^{+}$ [.200; 1.00]
Laplace			$.515^{+}$ (.185; .190)	$.970^{+}$	.297 (.406; .383)	$.923^{-}$	$.364^{-}$ [.230; .563]*	$.195^{-}$ [.085; .321]	$.573^{+}$ [.005; 1.00]
	70								
Normal		**10**	$.532^{+}$ (.276; .307)	$.980^{+}$	.283 (.345; .338)	.943	.462 [.263; .681]	.268 [.102; .524]	1.00^{+} [-.279; 1.00]
Laplace			.528 (.294; .313)	.960	.284 (.370; .348)	.953	$.562^{+}$ [.329; .843]*	.280 [.068; .547]	.988 [-.366; 1.00]
Normal		**40**	.503 (.147; .150)	.953	.329 (.260; .254)	$.930^{-}$	$.467^{-}$ [.352; .603]	$.231^{-}$ [.137; .348]	$.520^{+}$ [.192; 1.00]
Laplace			$.525^{+}$ (.149; .148)	.950	.322 (.250; .252)	.963	.492 [.351; .621]	.253 [.141; .366]	.422 [.069; .937]
Normal		**100**	.490 (.112; .104)	$.927^{-}$	.305 (.228; .221)	$.933^{-}$	$.472^{-}$ [.389; .554]	.247 [.177; .302]	.413 [.235; .606]
Laplace			$.511^{+}$ (.098; .103)	.957	.318 (.227; .218)	.947	$.453^{-}$ [.358; .571]	.236 [.164; .316]	$.346^{-}$ [.148; .529]*
True	**20**		**.500**		**.300**		**1.000**	**.500**	**.800**
Normal		**10**	$.632^{+}$ (.757; .762)	$.990^{+}$	.336 (.877; .842)	.940	$.869^{-}$ [.360; 1.522]	.553 [.117; 1.388]	1.00^{+} [-.541; 1.00]
Laplace			$.534^{+}$ (.706; .745)	$.983^{+}$	$.508^{+}$ (.921; .873)	.939	$.789^{-}$ [.315; 1.564]	.522 [.112; 1.187]	1.00^{+} [-.749; 1.00]
Normal		**40**	.499 (.376; .372)	$.967^{+}$	.245 (.673; .632)	.943	$.836^{-}$ [.509; 1.301]	.513 [.256; .870]	1.00^{+} [.764; 1.00]
Laplace			$.591^{+}$ (.348; .349)*	.957	.344 (.710; .624)	$.933^{-}$	$.832^{-}$ [.501; 1.233]	$.366^{-}$ [.148; .729]*	1.00^{+} [.627; 1.00]
Normal		**100**	.491 (.271; .259)	$.933^{-}$	.278 (.539; .528)	$.933^{-}$	$.861^{-}$ [.638; 1.155]	.483 [.258; .680]	1.00^{+} [.788; 1.00]
Laplace			$.541^{+}$ (.247; .241)*	.937	.337 (.576; .516)	$.920^{-}$	$.709^{-}$ [.495; 1.084]*	$.360^{-}$ [.194; .584]*	$.960^{+}$ [.597; 1.00]*
	70								
Normal		**10**	$.543^{+}$ (.342; .365)	$.977^{+}$	.338 (.452; .449)	.943	.935 [.646; 1.286]	.487 [.278; .823]	1.00^{+} [.602; 1.00]
Laplace			.498 (.381; .378)	$.977^{+}$	.300 (.487; .468)	.950	1.225^{+} [.751; 1.565]*	.471 [.186; .829]	1.00^{+} [.428; 1.00]
Normal		**40**	.509 (.183; .188)	$.967^{+}$	.303 (.355; .342)	.943	$.927^{-}$ [.766; 1.168]	.491 [.335; .626]	$.885^{+}$ [.657; 1.00]
Laplace			$.544^{+}$ (.186; .182)	.940	.377 (.353; .342)	.957	$.937^{-}$ [.694; 1.225]	$.438^{-}$ [.300; .598]*	$.719^{-}$ [.510; 1.00]
Normal		**100**	.496 (.148; .136)	$.927^{-}$	.303 (.289; .283)	$.933^{-}$	$.922^{-}$ [.786; 1.092]	.485 [.387; .588]	$.831^{+}$ [.719; .924]
Laplace			$.524^{+}$ (.128; .128)*	$.930^{-}$	.323 (.303; .290)	.940	$.896^{-}$ [.695; 1.101]	$.400^{-}$ [.290; .509]*	$.738^{-}$ [.586; .891]

* Significantly different from the correctly specified model.
$^{-}$ Significantly underestimated and $^{+}$ overestimated from the true value.

Table 5.8. Poisson modeling results for data generated from random coefficient Cox models with **50% censoring rate** and two different assumed random effects distributions: median of point estimates, empirical standard error and mean of the estimated standard error for fixed effects as well as probability coverage, and interquartiles ranges [$\mathbf{Q_1}$, $\mathbf{Q_3}$] for variance component estimates, over 300 simulated datasets.

True r.e.			β_1		β_2		σ_0^2	σ_1^2	ρ
dist'n	G	n_i	**mean (std; se)**	**Pr. C.**	**mean (std; se)**	**Pr. C.**	**median [Q_1; Q_3]**	**median [Q_1; Q_3]**	**median [Q_1; Q_3]**
True	**20**		**.500**		**.300**		**.500**	**.250**	**.400**
Normal		**10**	.514 (.260; .259)	$.967^{+}$	.316 (.490; .441)	$.920^{-}$	$.377^{-}$ [.229; .621]	$.190^{-}$ [.072; .462]	$.989^{+}$ [-.216; 1.00]
Laplace			.526 (.248; .253)	$.967^{+}$	.352 (.481; .423)	$.927^{-}$	$.372^{-}$ [.178; .618]	$.192^{-}$ [.060; .364]	.578 [-.489; 1.00]*
Normal		**40**	.477 (.150; .157)	.953	.247 (.415; .370)	$.920^{-}$	$.416^{-}$ [.299; .580]	$.216^{-}$ [.128; .317]	$.573^{+}$ [.216; .883]
Laplace			.508 (.154; .149)*	.937	.338 (.406; .357)*	$.927^{-}$	$.387^{-}$ [.247; .570]	$.169^{-}$ [.094; .282]*	$.483^{+}$ [.110; .905]
Normal		**100**	.503 (.133; .127)	$.923^{-}$	.301 (.365; .337)	$.933^{-}$	$.437^{-}$ [.317; .557]	$.210^{-}$ [.148; .301]	$.468^{+}$ [.200; .660]
Laplace			.509 (.119; .121)	.957	.260 (.363; .327)	$.927^{-}$	$.370^{-}$ [.246; .573]*	$.174^{-}$ [.115; .261]*	.410 [.187; .673]
	70								
Normal		**10**	.493 (.137; .136)	.946	.294 (.256; .244)	$.926^{-}$	$.457^{-}$ [.354; .563]	$.206^{-}$ [.122; .315]	$.597^{+}$ [.189; 1.00]
Laplace			.514 (.127; .132)*	.957	.264 (.258; .237)	$.920^{-}$	$.447^{-}$ [.332; .587]	$.167^{-}$ [.073; .309]*	.496 [.004; 1.00]
Normal		**40**	.492 (.082; .084)	.950	.306 (.205; .205)	.960	$.472^{-}$ [.400; .543]	$.230^{-}$ [.184; .288]	.427 [.278; .564]
Laplace			.506 (.081; .080)*	.940	.320 (.203; .199)	.950	$.435^{-}$ [.359; .539]*	$.193^{-}$ [.148; .248]*	$.348^{-}$ [.179; .552]*
Normal		**100**	.497 (.070; .070)	.943	.292 (.193; .187)	$.927^{-}$	$.464^{-}$ [.401; .528]	$.233^{-}$ [.195; .272]	.385 [.284; .488]
Laplace			.492 (.065; .067)	.963	.308 (.195; .185)	.937	$.444^{-}$ [.365; .535]	$.207^{-}$ [.172; .250]*	$.322^{-}$ [.209; .456]*
True	**20**		**.500**		**.300**		**1.000**	**.500**	**.800**
Normal		**10**	.524 (.293; .290)	$.970^{+}$	.269 (.644; .591)	$.933^{-}$	$.830^{-}$ [.546; 1.207]	$.303^{-}$ [.137; .562]	1.00^{+} [.578; 1.00]
Laplace			.538 (.275; .272)	.950	.336 (.634; .554)	$.903^{-}$	$.779^{-}$ [.472; 1.123]	$.258^{-}$ [.084; .453]	1.00^{+} [.145; 1.00]*
Normal		**40**	$.475^{-}$ (.190; .189)	.950	.302 (.502; .461)	$.933^{-}$	$.880^{-}$ [.638; 1.175]	$.365^{-}$ [.260; .499]	$.885^{+}$ [.666; 1.00]
Laplace			.506 (.179; .176)	.950	$.363^{+}$ (.512; .463)	$.930^{-}$	$.775^{-}$ [.541; 1.093]*	$.268^{-}$ [.172; .388]*	.814 [.500; 1.00]*
Normal		**100**	.493 (.172; .157)	$.926^{-}$	.338 (.439; .396)	$.920^{-}$	$.884^{-}$ [.690; 1.137]	$.349^{-}$ [.253; .466]	$.819^{+}$ [.672; .910]
Laplace			.485 (.142; .144)	.957	.297 (.462; .399)	$.920^{-}$	$.738^{-}$ [.514; 1.119]*	$.279^{-}$ [.194; .363]*	.767 [.582; .903]*
	70								
Normal		**10**	.494 (.156; .151)	.947	.291 (.330; .322)	.943	$.922^{-}$ [.780; 1.139]	$.318^{-}$ [.199; .462]	$.912^{+}$ [.657; 1.00]
Laplace			.512 (.142; .142)	.963	.273 (.334; .309)	$.930^{-}$	$.883^{-}$ [.689; 1.100]	$.227^{-}$ [.113; .355]*	.821 [.454; 1.00]
Normal		**40**	$.485^{-}$ (.097; .100)	.957	.306 (.249; .256)	.960	$.957^{-}$ [.832; 1.120]	$.368^{-}$ [.301; .450]	$.764^{-}$ [.669; .837]
Laplace			.501 (.089; .091)	.953	.327 (.266; .256)	.947	$.859^{-}$ [.731; 1.039]*	$.276^{-}$ [.221; .351]*	$.660^{-}$ [.503; .802]*
Normal		**100**	$.470^{-}$ (.091; .087)	$.923^{-}$	$.308^{+}$ (.220; .218)	.953	$.939^{-}$ [.823; 1.066]	$.391^{-}$ [.329; .446]	$.752^{-}$ [.692; .810]
Laplace			$.481^{-}$ (.076; .079)	.963	.328 (.235; .230)	$.933^{-}$	$.898^{-}$ [.740; 1.072]*	$.306^{-}$ [.257; .358]*	$.652^{-}$ [.557; .736]*

* Significantly different from the correctly specified model.
$^{-}$ Significantly underestimated and $^{+}$ overestimated from the true value.

Table 5.9. Poisson modeling results for data generated from nested random effects Cox models with **85% censoring rate** and two different assumed random effects distributions: median of point estimates, empirical standard error and mean of the estimated standard error for fixed effects as well as probability coverage, and interquartiles ranges [$\mathbf{Q_1}$, $\mathbf{Q_3}$] for variance component estimates, over 300 simulated datasets.

True r.e.				β_1		β_2		β_3		σ_1^2	σ_2^2
dist'n	n_i	$n_i G$	n_{ij}	**median (std; se)**	**Pr. C.**	**median (std; se)**	**Pr. C.**	**median (std; se)**	**Pr. C.**	**median [Q_1; Q_3]**	**median [Q_1; Q_3]**
True	**2**	**30**		**.500**		**.300**		**.200**		**.500**	**.250**
Normal			**10**	.553 (.323; .319)	.956	.178$^-$ (.786; .737)	.939	.219 (.768; .660)	.926$^-$	.358$^-$ [.102; .614]	.037$^-$ [.000; .302]
Log-gamma				.527 (.350; .328)	.963	.428 (.801; .726)*	.953	.150 (.714; .640)	.932$^-$	.322$^-$ [.067; .531]	.044$^-$ [.000; .263]
Normal			**40**	.497 (.161; .152)	.949	.159 (.596; .568)	.946	.218 (.518; .520)	.959	.456$^-$ [.316; .606]	.138$^-$ [.000; .273]
Log-gamma				.504 (.162; .157)	.946	.214 (.637; .573)	.915$^-$	.230 (.592; .526)	.912$^-$	.461$^-$ [.283; .641]	.114$^-$ [.000; .295]
Normal			**100**	.500 (.107; .095)	.916$^-$	.235 (.567; .520)	.916$^-$	.219 (.505; .482)	.940	.460$^-$ [.338; .613]	.124$^-$ [.000; .268]
Log-gamma				.506 (.097; .098)	.959	.260 (.589; .542)	.932$^-$	.209 (.556; .504)	.899$^-$	.512 [.378; .688]*	.124$^-$ [.000; .287]
	5	**75**									
Normal			**10**	.495 (.202; .197)	.952	.269 (.621; .506)	.864$^-$	.258$^+$ (.528; .457)	.881$^-$	.450$^-$ [.277; .609]	.132$^-$ [.033; .273]
Log-gamma				.491 (.195; .201)	.970$^+$	.388 (.564; .519)	.912$^-$	.194 (.536; .468)	.919$^-$	.449$^-$ [.314; .621]	.155$^-$ [.051; .289]
Normal			**40**	.493 (.095; .096)	.947	.338 (.464; .418)	.907$^-$	.162 (.443; .394)	.900$^-$	.484$^-$ [.402; .571]	.176$^-$ [.088; .279]
Log-gamma				.490 (.099; .099)	.960	.238 (.488; .432)*	.883$^-$	.264$^+$(.461; .407)*	.913$^-$	.518$^+$ [.434; .637]*	.182$^-$ [.093; .273]
Normal			**100**	.493 (.063; .060)	.936$^-$	.293 (.465; .388)	.900$^-$	.240 (.432; .370)	.896$^-$	.495 [.422; .575]	.165$^-$ [.086; .259]
Log-gamma				.507(.057; .062)*	.967$^+$	.258 (.429; .405)	.930$^-$	.218 (.411; .386)	.926$^-$	.550$^+$ [.473; .638]*	.160$^-$ [.088; .265]
True	**2**	**30**		**.500**		**.300**		**.200**		**1.000**	**.500**
Normal			**10**	.519 (.333; .334)	.969$^+$	.173 (.998; .940)	.948	.256 (.891; .848)	.938	.815$^-$ [.379; 1.224]	.157$^-$ [.000; .592]
Log-gamma				.514 (.314; .326)	.977$^+$	.293 (1.05; .934)	.930$^-$	.359$^+$ (.900; .840)	.943	.773$^-$ [.411; 1.216]	.162$^-$ [.000; .560]
Normal			**40**	.483 (.172; .161)	.963	.176 (.857; .778)	.937	.256 (.772; .715)	.937	.918$^-$ [.624; 1.210]	.243$^-$ [.000; .507]
Log-gamma				.479 (.164; .155)	.943	.313 (.877; .824)	.940	.250 (.908; .764)	.910$^-$	1.014 [.689; 1.438]*	.303$^-$ [.000; .654]
Normal			**100**	.492 (.104; .097)	.933$^-$	.251 (.811; .731)	.927$^-$	.206 (.709; .680)	.937	.934 [.682; 1.237]	.271$^-$ [.022; .530]
Log-gamma				.489 (.097; .098)	.970$^+$	.318 (.844; .803)	.926$^-$	.290 (.807; .747)	.930$^-$	1.132$^+$ [.875; 1.443]*	.313$^-$ [.000; .664]
	5	**75**									
Normal			**10**	.497 (.219; .205)	.940	.187 (.816; .665)	.880$^-$	.225 (.699; .610)	.903$^-$	.914$^-$ [.689; 1.223]	.286$^-$ [.100; .518]
Log-gamma				.488 (.199; .204)	.949	.272 (.782; .683)	.909$^-$	.117 (.704; .632)	.906$^-$	.957 [.726; 1.217]	.352$^-$ [.143; .590]*
Normal			**40**	.497 (.100; .098)	.953	.342 (.624; .573)	.916$^-$	.169 (.593; .545)	.912$^-$	.966 [.806; 1.152]	.341$^-$ [.171; .551]
Log-gamma				.502 (.097; .098)	.963	.331 (.717; .621)	.906$^-$	.180 (.674; .588)	.906$^-$	1.156$^+$ [.981; 1.356]*	.401$^-$ [.209; .607]
Normal			**100**	.495 (.063; .062)	.957	.288 (.656; .547)	.900$^-$	.227 (.608; .525)	.900$^-$	1.005 [.850; 1.179]	.337$^-$ [.180; .537]
Log-gamma				.503 (.064; .061)	.933$^-$	.236 (.707; .617)	.900$^-$	.152 (.643; .594)	.926$^-$	1.261$^+$ [1.035; 1.496]*	.443$^-$ [.228; .671]*

* Significantly different from the correctly specified model.

$^-$ Significantly underestimated and $^+$ overestimated from the true value.

Table 5.10. Poisson modeling results for data generated from nested random effects Cox models with **50% censoring rate** and two different assumed random effects distributions: median of point estimates, empirical standard error and mean of the estimated standard error for fixed effects as well as probability coverage, and interquartiles ranges [$\mathbf{Q_1}$, $\mathbf{Q_3}$] for variance component estimates, over 300 simulated datasets.

True r.e.				β_1		β_2		β_3		σ_1^2	σ_2^2
dist'n	n_i	$n_i G$	n_{ij}	**median (std; se)**	**Pr. C.**	**median (std; se)**	**Pr. C.**	**median (std; se)**	**Pr. C.**	**median [Q_1; Q_3]**	**median [Q_1; Q_3]**
True	**2**	**30**		**.500**		**.300**		**.200**		**.500**	**.250**
Normal			**10**	.503 (.182; .178)	.943	.224 (.593; .554)	.933$^-$	.224 (.558; .512)	.926$^-$	.406$^-$ [.249; .582]	.147$^-$ [.000; .324]
Log-gamma				.495 (.190; .175)	.920$^-$	.324 (.607; .571)	.920$^-$	.200 (.573; .526)	.913$^-$	.444$^-$ [.315; .639]	.134$^-$ [.000; .310]
Normal			**40**	.498 (.082; .084)	.963	.252 (.554; .500)	.930$^-$	.248 (.506; .466)	.910$^-$	.454$^-$ [.347; .577]	.136$^-$ [.000; .255]
Log-gamma				.500 (.080; .083)	.950	.273 (.557; .534)	.933$^-$	.227 (.533; .498)	.926$^-$	.547$^+$ [.386; .734]*	.147$^-$ [.000; .285]
Normal			**100**	.498 (.055; .053)	.938	.303 (.526; .487)	.942	.229 (.458; .456)	.935$^-$	.453$^-$ [.332; .605]	.128$^-$ [.017; .263]
Log-gamma				.493$^-$ (.052; .052)	.936$^-$	.294 (.568; .518)	.916$^-$	.144 (.545; .486)	.896$^-$	.548$^+$ [.405; .695]*	.134$^-$ [.005; .291]
	5	**75**									
Normal			**10**	.491 (.102; .111)	.967$^+$	.291 (.491; .410)	.887$^-$	.236 (.432; .386)	.923$^-$	.475$^-$ [.389; .564]	.149$^-$ [.073; .255]
Log-gamma				.490$^-$ (.109; .109)	.953	.291 (.486; .433)	.919$^-$	.225 (.480; .410)	.886$^-$	.534$^+$ [.442; .633]*	.184$^-$ [.098; .296]*
Normal			**40**	.495 (.054; .053)	.949	.339 (.418; .381)	.909$^-$	.162 (.409; .367)	.889$^-$	.491 [.415; .562]	.171$^-$ [.099; .265]
Log-gamma				.491$^-$ (.054; .053)	.933$^-$	.298 (.457; .417)	.919$^-$	.231 (.435; .400)*	.902$^-$	.595$^+$ [.502; .699]*	.188$^-$ [.106; .277]
Normal			**100**	.492$^-$ (.032; .033)	.950	.303 (.423; .372)	.907$^-$	.217 (.408; .359)	.880$^-$	.487 [.424; .562]	.173$^-$ [.095; .263]
Log-gamma				.498 (.031; .033)	.960	.269 (.425; .396)	.919$^-$	.215 (.401; .380)	.933$^-$	.605$^+$ [.514; .699]*	.170$^-$ [.090; .267]
True	**2**	**30**		**.500**		**.300**		**.200**		**1.000**	**.500**
Normal			**10**	.499 (.180; .180)	.956	.186 (.795; .731)	.922$^-$	.227 (.740; .682)	.929$^-$	.813$^-$ [.545; 1.108]	.277$^-$ [.000; .610]
Log-gamma				.508 (.174; .181)	.967$^+$	.202 (.914; .814)	.933$^-$	.297$^+$ (.803; .753)	.930$^-$	1.112$^+$[.752; 1.554]*	.250$^-$ [.000; .570]
Normal			**40**	.486$^-$ (.081; .084)	.960	.187 (.773; .692)	.923$^-$	.311 (.708; .646)	.920$^-$	.881$^-$ [.700; 1.153]	.271$^-$ [.020; .506]
Log-gamma				.479$^-$ (.086; .085)	.959	.295 (.822; .799)	.943	.285 (.842; .749)	.912$^-$	1.228$^+$ [.914; 1.560]*	.349$^-$ [.000; .704]
Normal			**100**	.491$^-$ (.056; .053)	.926$^-$	.289 (.711; .684)	.946	.225 (.626; .642)	.946	.876$^-$ [.675; 1.188]	.262$^-$ [.042; .520]
Log-gamma				.487$^-$ (.055; .054)	.943	.294 (.851; .801)	.926$^-$	.195 (.802; .750)	.926$^-$	1.302$^+$ [.992; 1.660]*	.308$^-$ [.000; .696]
	5	**75**									
Normal			**10**	.480$^-$ (.105; .113)	.966$^+$	.271 (.668; .551)	.895$^-$	.234 (.585; .525)	.926$^-$	.932$^-$ [.754; 1.090]	.315$^-$ [.183; .484]
Log-gamma				.497 (.111; .114)	.963	.321 (.694; .615)	.906$^-$	.108$^-$ (.645; .588)*	.900$^-$	1.213$^+$ [1.030; 1.420]*	.397$^-$ [.188; .646]*
Normal			**40**	.483$^-$ (.053; .054)	.950	.326 (.578; .529)	.906$^-$	.149 (.565; .512)	.893$^-$	.948$^-$ [.808; 1.103]	.341$^-$ [.202; .522]
Log-gamma				.487$^-$ (.053; .054)	.950	.322 (.683; .608)	.906$^-$	.158 (.650; .583)	.926$^-$	1.341$^+$ [1.154; 1.544]*	.420$^-$ [.206; .646]*
Normal			**100**	.484$^-$ (.034; .034)	.913$^-$	.295 (.600; .516)	.907$^-$	.251 (.577; .499)	.883$^-$	.953$^-$ [.841; 1.100]	.328$^-$ [.190; .491]
Log-gamma				.485$^-$ (.035; .034)	.910$^-$	.200 (.719; .619)	.896$^-$	.157 (.661; .597)	.920$^-$	1.458$^+$ [1.243; 1.711]*	.440$^-$ [.234; .667]*

* Significantly different from the correctly specified model.

$^-$ Significantly underestimated and $^+$ overestimated from the true value.

Chapter 6

Concluding remarks and future research

6.1. Introduction

The main objectives of this research project were twofold: first, to explore parameter estimation for semiparametric hierarchical proportional hazards models beyond the simple (shared) frailty models, and to evaluate through simulations the performance of the existing estimation methods for these models. The performance of the methods were investigated in terms of statistical properties such as unbiasedness, robustness, and confidence interval coverage. Second, two animal health datasets with thousands of records and time-dependent predictors and effects were analyzed.

The specific objectives for each study involved in the thesis were as follows. The objective of the first study was to establish some practical guidelines for the choice of appropriate estimation procedures for estimating Cox models with random intercept and random slope (Chapter 2). The second study aimed to analyze a 3-level veterinary dataset with time-dependent predictors and coefficients, and to evaluate the performance of the approach used and compare it with other existing approaches (Chapter 3). The purpose of the third study was to demonstrate the use of the Poisson maximum likelihood approach, in concert with posterior Bayesian inference, for estimating a cross-classified and multiple membership model with time-to-event response. Further, the multiple membership model analysis was applied to a large dataset from veterinary medicine (Chapter 4). Finally, the objective of the fourth study was to assess the robustness of Poisson maximum likelihood estimation for Cox models with normal random effects under misspecification of the random effects distribution (Chapter 5).

In this chapter, the main findings and contributions of this work are summarized. Some future perspectives are also briefly discussed.

6.2. Multi-component frailty Cox models

6.2.1. Cox model with random intercept and random coefficient

In Chapter 2, through simulations, the performance of four procedures for estimating Cox models with random intercept and random slope commonly used in epidemiology and implemented in broadly accessible statistical software was compared. These procedures were: the penalized partial likelihood (Ripatti and Palmgren, 2000; coxme), the penalized full likelihood (Rondeau et al., 2008; frailtypack), the Poisson maximum likelihood (Ma et al., 2003; Feng et al., 2005; xtmepoisson), and the Bayesian approach (WinBUGS). The simulation was designed to mirror settings of real animal health data. For each simulated dataset, two random intercept and random slope Cox models were fit; one assumed a diagonal covariance matrix for random effects and the other used a full covariance matrix.

In this study, convergence problems were observed for PFL and PML, but the non-convergence rate for the latter was lower and can be dealt with by changing the integration points of the Gauss-Hermite method. The PFL approach exhibited very low convergence rates when the magnitude of variance components was large and the optimum convergence rates were not assured even for a small magnitude of heterogeneity. When the covariance structure for random effects was correctly specified, the study showed satisfactory results for both independent and correlated random effects model analyses, whereas ignoring existing correlation between random effects led to biased estimates for fixed effect parameters when the variance was large, and such bias increased with increasing σ_{01}.

Moreover, results from this comparative study showed that in terms of estimation of the fixed effect parameter, all the estimation procedures yielded good and comparable estimates when the between-cluster variability for the random effects was limited. In contrast, when the magnitude of variability for the random effects was large, only the PML and BAY procedures produced reasonable estimates for the fixed effect parameter, while it was underestimated by

PPL and the estimates of PFL were not of interest due to the high rate of non-convergence. The estimated and empirical SEs agreed closely for the PML, PFL, and BAY procedures. However, PPL underestimated the SEs of fixed effect estimates when $\sigma_{01} \neq 0$. This underestimation of fixed effect SE was also pointed out in Ripatti and Palmgren (2000) and Therneau and Grambsch (2000, p. 249). For random effect parameter estimation, the performance of all the procedures for estimating σ_0^2 was generally good, one exception being that PPL underestimated variance when the variance components were large. With the exception of PML with large variances, all the procedures produced estimates for σ_1^2 that were mostly biased upwards when the variance components were small to moderate, and somewhat biased downwards for a large magnitude of variance components. All the estimation procedures tended to estimate σ_{01} with downwards bias in most cases. The SEs for σ_0^2, σ_1^2, and σ_{01} were estimated with reasonable accuracy by PML, and by PFL when variance components were small to moderate, whereas the mean of posterior standard deviation for BAY overestimated such SEs.

In conclusion, estimating the covariance between the random intercept and the random slope in the analysis is recommended and the PML approach seems to be a preferable choice for this task.

6.2.2. Cox model with two nested frailties

The current implementation of nested frailty Cox models, such as those implemented in coxme, frailtyHL, and frailtypack packages of R software, are limited to models with few predictors and moderate size of datasets. The coxme and the frailtyHL implementations of nested frailty models do not support the counting process format necessary for modeling time-dependent predictors. On the other hand, the implementation of nested frailty models in the frailtypack package assumes a gamma distribution for frailties and allows for time-

dependent predictors and coefficients, but it requires the number of cluster to be at most moderate (a dataset of 6973 subjects from 18 areas nested in 6 cities was used in the original paper of Rondeau et al. (2006)). In Chapter 3, we applied a nested log-normal frailty model to a large dataset from a veterinary epidemiologic field with quite a large number of independent variables (many parameters); these independent variables include time-varying predictors and effects. The model was fit to the data as a Poisson GLMM after transforming the data to the counting-process format and use of the Gaussian quadrature method for a more accurate ML estimation. The estimation approach was then evaluated via a simulation study with a data structure similar to the real data. Through the simulation study, the performance of PMLGQ, PMLAP and PPL was compared in terms of bias in point estimates and their SEs. The study showed that the approach used with Gaussian quadrature produced fairly robust and adequate estimation for both fixed and random effect parameters as well as their SEs. On the other hand, the approach with Laplace approximation produced estimates for the between-subcluster variance with strongly upwards bias and underestimated the between-cluster variance and the SEs of variance components. In comparison, the performance of PPL was good and comparable to PMLGQ in estimating the fixed effect parameters and their SEs. However, the procedure underestimated the between-subcluster variance. The take-home message from this comparison was that PMLGQ performed best.

6.2.3. Cox model with a multiple membership and cross-classified frailties

In this thesis, a cross-classified and multiple membership model was applied to time-to-event data. To the best of our knowledge, such a model has not previously been reported in the survival analysis literature. The model was estimated, through the identity between the Cox PH model and the Poisson model, as a mixed-effects Poisson model using Bayesian techniques. A simulation study was conducted to evaluate the approach using settings

resembling real data and another setting with a larger magnitude of heterogeneity and a more pronounced data structure. Results indicated that the proposed estimation approach performed well in estimating most of the model parameters when the data structure and the variance components were similar to the real data (8% of herds in a multiple membership and small variance components). Further, the estimation was clearly improved in the settings of a large magnitude of variance components and a more pronounced CMM data structure.

6.2.4. Misspecification of the frailty distribution

A simulation study was conducted to assess the robustness of Poisson maximum likelihood estimation for Cox models with normally distributed random effects against misspecification of the random effects distribution. The consequences of assuming normality of random effects to distributions that are clearly non-normal were examined. The impact of misspecification was assessed in three different frailty model designs, namely, shared frailty, random coefficient, and nested frailty Cox models. The results showed that the approach used provides robust estimation for fixed effects even with misspecified random effects distribution. On the other hand, misspecification of the random-effects distribution may have substantial impact on the estimation of variance components, especially when the magnitude of variability between random effects, number of groups, group size, and event rate are large. In these situations, caution needs to be exercised in the interpretation of the analysis.

6.3. Analyses of real datasets

In Chapter 2, a subset of the lameness data (Christensen, 1996) was used as an example to study the performance of four estimation methods based on a random coefficient Cox model. A Cox model with random herd and treatment effects was applied to the dataset and heterogeneity in the baseline hazard and in treatment effects between herds was quantified using the four estimation methods. Two analyses were carried out; one assumed a diagonal

covariance matrix for normally distributed random effects and the other used a full covariance matrix. Both analyses demonstrated substantial variation in the baseline hazard and in the treatment effect between herds, as well as a clear discrepancy in the estimation of fixed treatment effect between the two analyses. The simulation findings concur with such discrepancy and it may be attributable to the misspecifying of the correlation structure between random effects. Analysis and further exploration of the full lameness data can be found in (Stryhn and Christensen, 2013).

In Chapter 3, a dataset from the Canadian Bovine Mastitis Research Network was analyzed to identify and evaluate risk factors associated with the hazard of clinical mastitis. The CBMRN data had 8,035 cows clustered by 69 herds; some of these cows with multiple lactations resulted in a dataset with two hierarchical levels of clustering at cow and herd levels. Due to the presence of time-varying predictors and effects, large number of predictors, large number of random effects at the cow level as well as the size of the CBMRN dataset, none of the existing frailty model software was able to handle such a dataset. Carrying over the identity between the Cox PH model and a Poisson model to the nested frailty models and using the theory and software for GLMMs, allowed us to fit a log-normal nested frailty Cox model to the CBMRN dataset. However, using this estimation approach can be challenging for such a large dataset as the data expansion into the counting-process format led to a massive dataset, that complicated the analysis and required excessive computing time. Moreover, the large number of predictors in such large data produced complex interactions between predictor effects and increased the chance of violating model assumptions such as the proportionality of hazards, as well as complicating the model building process and the interpretation of results. In conclusion, the analysis of the CBMRN dataset demonstrated substantial variation in the baseline hazard among cows and between herds, and also indicated that some of the herd managerial factors combined with cow characteristics influenced the hazard of CM

during the lactation period; most of these effects had a strong effect early rather than later in the lactation. There was great value in accounting for the full structure of the CBMRN dataset by estimating variance components at the cow and herd levels and at the same time modeling time-varying predictors and effects. The PML approach offers the opportunity to analyze such large and complex datasets and models. However, the low incidence of failures and the large number of coefficients made model convergence challenging to achieve. Changing the integration points of the Gaussian quadrature method may help in these cases. Finally, the analysis of large datasets with quite many and complex predictors, such as in the CBMRN dataset, presents challenges beyond the estimation in itself. Such analyses may result in models with a complex network of interacted effects, e.g. including interaction with time. These effects are often hard to present and interpret. In this regard, the graphical presentation of predictor effects as well as purposely selected hazard ratios may aid the interpretations.

In Chapter 4, a survival analysis was carried out for a dataset from the Western Canada beef productivity study (Waldner, 2008) to examine individual, herd management, and environmental factors associated with calf loss and mortality in beef cattle. The dataset was large in size and had a special structure. In addition to calves being hierarchically nested within herds, herds were cross-classified by veterinary clinics and ecological regions, and some of the herds were serviced by two veterinary clinics, creating a 3-level CMM data structure. A CMM frailty model was fit to the calf mortality dataset to account for both the special data structure and the time-varying effects for some predictors. The model was fit as a CMM Poisson model and estimated using Bayesian techniques which were computationally demanding. The analysis of calf mortality data showed a larger variation between herds than between both veterinary clinics and ecological regions, and a clear improvement in model fit after accounting for the variation between veterinary clinics and ecological regions. Furthermore, the analysis demonstrated that some of the individual and environmental factors

as well as some predictors related to biosecurity practices influenced the hazard of mortality in calves. Finally, we conclude from this and previous work that there was a benefit in adding complexity to a simple frailty model by taking into account both the time-varying effects and the hierarchical data structure. In addition to the improvement in the model fit, variances for herd, veterinary clinic, and ecologic region were quantified. MLwiN software provides the facilities to estimate the model as a CMM Poisson model using MCMC techniques. However, the analysis required excessive computing time.

6.4. Estimation approaches

6.4.1. Poisson modeling approach for frailty models

The equivalence of a Cox proportional hazards model with a Poisson regression model (PML) has been known since the 1980s (Whitehead, 1980). Such equivalence can be carried over to the case of hierarchical survival data (Ma et al., 2003; Feng et al., 2005). Therefore, frailty models with an unspecified (or a piecewise constant) hazard function can be estimated as mixed-effects Poisson models. Using GLMM software, we can fit frailty models with random coefficients or several hierarchical levels of clustering as well as fitting frailty models to data with imperfect hierarchical structure such as cross-classified levels or/and multiple membership structure. Furthermore, with the Poisson modeling approach we can apply the adaptive Gaussian quadrature method for accurate maximum likelihood estimation or use Bayesian techniques for estimation. As the data need to be expanded to apply the approach, the latest version of Stata, where the memory is resized automatically, provides opportunity to analyze very large datasets (millions of records) with many predictors including time-varying predictors but of course with compromises in computing time. Also, Stata facilities for Poisson GLMMs are readily available for inference.

The Poisson modeling approach showed, in general, good performance in estimating different designs of frailty models including shared frailty, random coefficient, and nested frailty models using maximum likelihood estimation, and a cross-classified and multiple membership frailty model based on MCMC estimation. However, the approach tended to produce unrealistic SEs and CIs for maximum likelihood estimates of variance components when the number and the size of clusters as well as the event rate were small. The simulation study in Chapter 5 showed that the PML approach produced robust estimates under misspecification of the random effects distribution for both the within-group and between-group fixed effects. Further, the approach also gave robust estimates for variance components in a wide variety of commonly encountered situations in veterinary medicine when both the magnitude of heterogeneity and event rate were small. Misspecification of the random effects distribution may become a problem for variance component estimates when the magnitude of heterogeneity, number of groups, group sizes, and event rate are large.

6.4.2. Penalized partial likelihood approach

The performance of the PPL approach (Ripatti and Palmgren, 2000) was assessed in this thesis based on a random coefficient Cox model including one individual-level predictor and simulation settings with varying magnitude of heterogeneity and heavy censoring. The approach was also evaluated based on a Cox model with two normally distributed nested random effects including three predictors at different hierarchical levels and simulated datasets of correlated event times with a high rate of censoring. The approach performed well in estimating the fixed effect parameters and SEs under the nested frailty model and the random coefficient model as long as the magnitude of heterogeneity was small. For a large magnitude of heterogeneity, PPL underestimated the within-cluster fixed effect as well as its SE when a full covariance matrix for random effects was assumed in the analysis. Based on

the random coefficient model, the PPL approach performed reasonably well in estimating the variance components for a relatively small magnitude of heterogeneity and tended to underestimate them for a large magnitude of heterogeneity. Similarly, PPL gave a good estimate for the variance at the cluster level in nested frailty model and underestimated it at the subcluster level. Finally, the current implementation of PPL in R software (coxme) did not provide SEs for variance component estimates nor support the counting-process data format needed for modeling time-varying predictors and effects. However, its algorithm was fast to converge and did not suffer from convergence problems.

6.4.3. Other estimation approaches

Two other estimation approaches were used in this thesis; the penalized full likelihood (Rondeau et al., 2008) and Bayesian approaches. The performance of both approaches was assessed based on a random coefficient Cox model, so our discussion will be limited to this model design. The PFL approach experienced a lot of convergence difficulties, especially in the cases of a large magnitude of variance components, such a convergence problem was also pointed out in Hirsch and Wienke (2012) but based on a shared frailty model. PFL performed reasonably well in estimating all model parameters and their SEs as long as the magnitude of heterogeneity was small to moderate. The Bayesian approach, on the other hand, showed a good performance in estimating the fixed effect and its SE as well as the variance of random intercept, whereas it gave estimates for the random slope variance biased upwards for a relatively small magnitude of heterogeneity and biased downwards for a large magnitude of heterogeneity. Further, the posterior standard deviations tended to overestimate the variability between variance component estimates. Finally, the Bayesian procedure was a computer-intensive technique and required special software.

6.4.4. Recommendations

The multi-component frailty models including the random coefficient model, nested, and non-nested frailty models have several advantages over one-component (shared) frailty models. In addition to quantifying heterogeneity in the baseline hazard at different hierarchical levels, they improve the model fit and the estimation of fixed effects and handle predictors at different hierarchical levels as well as allowing for predictor effects to vary between clusters. Stumbling blocks in the use of multi-component frailty models relate to the estimation of parameters and the lack of software. One recommendation to overcome these problems is to convert frailty models into mixed-effects Poisson models and generate estimations using available software for GLMMs. Other specific recommendations that can be drawn from this work are:

(1) The use of the Poisson modeling approach in concert with posterior Bayesian inference is recommended for modeling imperfectly hierarchical survival data, although further research is needed to confirm the validity of this approach under different circumstances.

(2) The Poisson maximum likelihood approach with Gaussian quadrature is always recommended. However, for a small magnitude of heterogeneity unrealistic SEs and CIs may be obtained by the approach based on shared and nested frailty models when the number of clusters, cluster size, and event rate are small; in such cases these statistics should not be used for inference.

(3) For substantial heterogeneity, the dataset has to be carefully inspected and if possible different frailty distributions have to be used in the analysis to check the robustness of the model parameter estimates.

(4) In random coefficient models, estimating the correlation between the random intercept and the random slope is recommended.

(5) The penalized partial likelihood approach is recommended for quick data exploration and its fixed effects estimates may be used for inference as long as the magnitude of heterogeneity in the data is small.

6.5. Future perspectives

Although, the thesis presents different model designs and their estimation approaches for analyzing hierarchical survival data beyond the shared frailty model analysis, there are many other aspects related to the analysis of hierarchical survival data that are not covered and may be considered as topics for future research. In Chapters 2 and 3, the simulation studies were built to mirror settings of specific datasets and the performance of the estimation procedures used was evaluated based on those settings. Further research may be needed to assess the effect of other factors such as the censoring rate, number of clusters, cluster size, type of predictor and type of baseline hazard on the performance of estimation procedures. Furthermore, evaluating the feasibility of 3-level Cox models with random coefficients using either the Poisson modeling approach or extending one of the other existing estimation approaches may be of great interest in the future. Another thought is to explore and apply time-dependent frailty models that can deal with the situation where we believe that the time-varying frailty exists; this phenomenon might be found in large veterinary data. Moreover, development of likelihood-based estimation techniques for special frailty models such as multiple membership and cross-classified frailties could also be a topic for future research. A possible solution might be to use GLLAMM and implement the model as a mixed-effects Poisson model. There are other open issues such as, extending frailty model software to support counting-process formatted data, providing tools for testing proportional hazards and goodness-of-fit in frailty models and quantifying correlation at different hierarchical levels, developing an algorithm for simulating survival data from a frailty model with time-varying

effects, and extending other survival models to allow for different data structures in the cases where PH models do not provide adequate analysis.

References

Christensen J. (1996). The influence of selected litter and herd factors on treatments for lameness in suckling piglets from 35 Danish herds. *Preventive Veterinary Medicine*, **26**: 107-118.

Feng S, Wolfe R, Port F. (2005). Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *Journal of the American Statistical Association*, **100**: 718-735.

Hirsch K, Wienke A. (2012). Software for semiparametric shared gamma and log-normal frailty models: An overview. *Computer Methods and Programs in Biomedicine*, **107**: 582-597.

Ma R, Krewski D, Burnett R, (2003). Random effects Cox models: A Poisson modelling approach. *Biometrika*, **90**: 157-169.

Ripatti S, Palmgren J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, **56**: 1016-1022.

Rondeau V, Filleul L, Joly P. (2006). Nested frailty models using maximum penalized likelihood estimation. *Statistics in Medicine*, **25**: 4036-4052.

Rondeau V, Michiels S, Liquet B, Pignon J. (2008). Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine*, **27**: 1894-1910.

Stryhn H, Christensen J. (2014). The analysis – hierarchical models: past, present and future. *Preventive Veterinary Medicine*, **113**: 304-312.

Therneau T, Grambsch P. (2000). Modelling Survival Data, Extending the Cox model. New York: Springer.

Waldner C. (2008). Western Canada study of animal health effects associated with exposure to emissions from oil and natural gas field facilities. Study design and data collection I. Herd performance records and management. *Archives of Environmental & Occupational Health No 4*, **63**: 167-184.

Whitehead J. (1980). Fitting Cox's regression model to survival data using GLIM. *Journal of Applied Statistics*, **29**: 268-275.