NONPARAMETRIC FRAILTY MODELS FOR CLUSTERED SURVIVAL DATA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Emmanuel Sebastian Sharef Moreno

August 2008

NONPARAMETRIC FRAILTY MODELS FOR CLUSTERED SURVIVAL DATA

Emmanuel Sebastian Sharef Moreno, Ph.D.

Cornell University 2008

The dependence between subjects in clustered survival data is commonly modeled by means of a frailty, a multiplicative random effect with a distribution that is usually specified in advance. Misspecification of the frailty distribution can lead to error when estimating parameters of interest.

This dissertation contains two distinct approaches to frailty models for the analysis of clustered survival data that do not require the frailty distribution to be known a priori.

The first is a Bayesian method, in which the distribution of both the baseline hazard and frailty are modeled nonparametrically as mixtures of B-splines, and estimated by Markov Chain Monte Carlo. Smooth curve estimates are attained by adaptive selection of the spline knots, or by means of an explicit smoothness penalty. The method is illustrated with data sets from studies of congestive heart failure and diabetic retinopathy.

The second is a method for clustered bivariate recurrent event data, in which the hierarchical bivariate frailty need only be specified through its first two moments. Estimation relies on a correspondence between the modulated renewal process likelihood and an auxiliary Poisson model likelihood, which allows the frailties to be estimated by their best linear unbiased predictors in an iterative algorithm. Data on recurrent basal and squamous cell carcinomas collected during the Nutritional Prevention of Cancer trial serves to illustrate the method.

BIOGRAPHICAL SKETCH

Emmanuel Sharef was born on November 5, 1981, in Cali, Colombia, to Uriel Sharef and Ana Sofia Moreno de Sharef.

He completed primary education in Colombia and Germany, attending respectively the Colegio Andino in Bogotá, and the Hans–Sachs–Gymnasium in Nürnberg.

In 2003, he earned an Bachelor of Science in Engineering from the department of Operations Research and Financial Engineering at Princeton University, under the supervision of Damir Filipović and Erhan Çınlar.

He carried out his doctoral research at the department of Operations Research and Information Engineering at Cornell University, advised by David Ruppert, Robert Strawderman, and Philip Protter.

Emmanuel now begins a career in the financial industry.

ACKNOWLEDGEMENTS

I am grateful to the following:

My parents, Uriel Sharef and Ana Sofia Moreno de Sharef, and my sister Eleonora Sharef, for their continuous support. My grandmothers, Nina Sharef and Ifigenia Jaramillo de Moreno, and my aunts, uncles and cousins.

My special committee co-chairs, David Ruppert and Rob Strawderman, and my minor advisor Philip Protter, for their mentoring, brainstorming, problem–solving and editorial assistance. Their guidance and patience over the years has been invaluable, and has informed and substantially improved the research associated with this thesis.

The faculty of the ORIE department and Cornell University, particularly Bruce Turnbull, Stefan Weber, Robert Jarrow, Antje Berndt, Narahari Prabhu, Huseyin Topaloglu, Gennady Samorodnitsky, Adrian Lewis, and Qiuyun Teng.

The staff of the department of ORIE and the Cornell Graduate School.

For their help with data analysis and interpretation, James Marshall, Mary Reid, and Raj Natarajan of the Roswell Park Cancer Institute, and Mark Cowen and Lakshmi Halasyamani of St. Joseph Mercy Hospital.

My officemates at Cornell, Jie Chen, Yun Shi, Anke van Zuylen, and Stefan Wild, as well as my fellow students in the department.

Old friends, especially Steven Aller, Lance Baird, Michael Kimberly, Jon Ophardt, Jim Hunt, Khalil Sullivan, Ryan Peterson, Yu Yu, and Hannah Chang.

Thank you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CLUSTERED SURVIVAL AND RECURRENT EVENT DATA

Survival analysis is concerned with the study of time-to-event data. A data set consists of covariate information describing a number of subjects, each of which is monitored until an event of interest occurs, or the subject is removed from the sample for an unrelated reason. The objective is to quantify the possibly time-varying risk that a subject experiences such an event, taking into account all available information. In typical model formulations, this takes the form of estimating a baseline risk shared by all subjects, and quantifying the effect of subject-level covariates on this risk. It is common that several subjects do not experience an event during the period of observation. This phenomenon is known as censoring, and methods for survival analysis must be designed to yield valid results regardless of its presence.

For example, Chapter 2 considers a data set of diabetic retinopathy patients enrolled in a six-year clinical trial of the effectiveness of laser photocoagulation treatment in preventing blindness. In this study, each one of the patients' eyes may be treated as a "subject", the event of interest is the onset of blindness, and covariates indicate whether the eye has received treatment or a placebo, as well as the type of diabetes and other patient data. A researcher may wish to quantify the risk of blindness at different times for treated and untreated eyes, and determine whether any difference is statistically significant, while properly accounting for other distinctions between the patients, and correctly including patients who did not become blind during the study in the analysis.

Applications outside of the medical arena are plentiful as well, for example the analysis of corporate bankruptcy risk, or reliability studies for manufactured

components.

The proportional hazards model introduced by Cox (1972a) remains by far the most widely used method for survival analysis in the statistical and medical literature. The model assumes that at-risk subjects in a sample of size J experience events at times $T_j\,, j = 1 \ldots J$, which are independent realizations of single-jump counting processes with intensity (often also called hazard rate):

$$\lambda_j(t|\boldsymbol{Z}_j) = \lambda_0(t)e^{\boldsymbol{\beta}^T\boldsymbol{Z}_j}\ , \quad j = 1 \ldots J\,. \tag{1.1}$$

Here, $\lambda_0(t)$ is known as the baseline hazard shared among all subjects, $\boldsymbol{Z}_j$ is a vector of covariates for subject j, and $\boldsymbol{\beta}$ is a vector of regression coefficients. In the proportional hazards formulation, the coefficients $\boldsymbol{\beta}$ capture the effect of the covariates on the event risk, and are therefore the objects of ultimate interest in most analyses. The proportional hazards model is named thus because the ratio of hazards for two subjects does not depend on the baseline hazard, that is, $\lambda_1(t|\boldsymbol{Z}_1)/\lambda_2(t|\boldsymbol{Z}_2) = \exp(\boldsymbol{\beta}^T(\boldsymbol{Z}_1 - \boldsymbol{Z}_2))$, allowing the intuitive interpretation of the estimated regression parameters $\boldsymbol{\beta}$ as the effect of a change in covariate $\boldsymbol{Z}$ on the log-hazard rate.

Thanks to an ingenious partial likelihood approach, proposed in Cox (1972a) and formalized in Cox (1975), the coefficients $\boldsymbol{\beta}$ can be estimated separately from the baseline hazard $\lambda_0(t)$, resulting in a quick and accurate method for simple survival data sets. The methodology can be extended to data with tied event times, time-dependent covariates, strata and different censoring mechanisms with relative ease. Andersen et al. (1993) present the Cox model and its extensions in detail, and Kalbfleisch and Prentice (2002) give a thorough discussion of the procedure's theoretical underpinnings. Therneau and Grambsch (2000) present current software implementations in `R` and `SAS` and discuss their proper use.

For many survival data sets, the assumption that events occur independently of each other is a reasonable one, particularly if nothing is known about the relationship between subjects. Notable exceptions are clustered data and recurrent event data. These two cases are the focus of this dissertation, and will be briefly presented in the sections that follow.

1.1 Clustered survival data

In real-world data sets, subjects are often grouped into "clusters" that share potentially unmeasurable similarities. For instance, clinical trials may involve large numbers of clinics or physicians; subjects treated by the same physician generally live in geographical proximity, may come from a similar socioeconomic background and ideally receive the same quality of care. Such forms of dependence between subjects cannot be reliably captured by means of covariates, yet ignoring it entirely can lead to unacceptable errors.

In the aforementioned diabetic retinopathy study for example, treating each eye as an independent subject would be inappropriate, as every patients' two eyes share the same genetic makeup and life-history. Such factors cannot be well-captured by covariates, but may have even greater influence on the risk of blindness than the treatment itself. If the effectiveness of treatment is to be accurately assessed, the unmeasured similarity between a patients' eyes must be addressed.

Frailty models are an extension of the Cox model that allows such dependence to be captured in an intuitive way. In the shared frailty approach, a multiplicative random effect is common to all members of a cluster. That is, for a sample of m clusters of size J_i, $i = 1 \dots m$, the event times T_{ij} are independent conditionally

on a set of frailties U_i, $i = 1 \ldots m$, so that the intensity of eq. (1.1) is replaced by:

$$\lambda_{ij}(t|\boldsymbol{Z}_{ij}, U_i) = U_i \lambda_0(t) e^{\boldsymbol{\beta}^T \boldsymbol{Z}_{ij}} \,, \quad i = 1 \ldots m, \quad j = 1 \ldots J_i \,, \tag{1.2}$$

for at-risk subjects, where the frailties U_i, $i = 1 \ldots m$ are independent and identically distributed from some predetermined distribution with mean 1. Intuitively then, the hazard rate for subject (i, j) is composed of (1) the baseline hazard shared by all subjects, (2) the frailty multiplier shared by subjects in cluster i, and (3) the adjustment for the covariates of subject (i, j). Each of these components must be estimated in order to quantify the risk for subject (i, j). Fixing the frailty mean at 1 ensures that the baseline hazard can be identified.

Even though the hazard specifications of eqns. (1.1) and (1.2) are similar, the partial likelihood approach of Cox (1975) is no longer feasible for frailty models, because the proportional hazards assumption does not hold for the marginal hazards—that is, unconditionally on the frailties. It is therefore necessary to estimate the regression parameters $\boldsymbol{\beta}$, the baseline hazard $\lambda_0(t)$, the frailties $\boldsymbol{U}$ and any additional parameters required by one's model specification simultaneously.

In the frailty model originally proposed by Clayton and Cuzick (1985), the frailties U_i are assumed to arise from a gamma distribution, and the baseline hazard $\lambda_0(t)$ is parametrically specified as either an exponential or Weibull hazard. Under these strong assumptions, one can compute and maximize the marginal likelihood of the unknown parameters, including the regression coefficients $\boldsymbol{\beta}$. The NPMLE (nonparametric maximum likelihood estimation) approach of Nielsen et al. (1992) allows the baseline hazard to be specified nonparametrically, as a Breslow-type stepfunction estimator, but retains the gamma frailty assumption; the model is fit by an iterative Expectation-Maximization (EM) algorithm, which alternates between estimating the frailties by their best unbiased predictors (BUPs) and

maximizing the joint likelihood of the regression parameters and baseline hazard. Assuming gamma frailties allows the BUPs to be computed in closed form, reducing computational cost and making the asymptotics tractable (Parner, 1998).

Other parametrized distributions for the frailties have been proposed, with corresponding variations on the gamma NPMLE, including lognormal (McGilchrist and Aisbett, 1991) and stable distributions (Hougaard, 1986). Models with gamma frailties are equivalent to penalized Cox models (Ripatti and Palmgren, 2000; Therneau and Grambsch, 2000), and can be more easily estimated in this way, as implemented by the `coxph` procedure in `R`. Similarly, lognormal frailty models can be well-approximated by penalized models.

The aforementioned frailty methods suffer from the need to specify the distribution of the frailty parametrically. When conducting data analysis, there is in fact rarely an a priori reason to choose one frailty distribution over another; rather, the choice is often dictated by the availability and capability of model-fitting software. When the chosen frailty distribution poorly approximates the distribution of unobservable random effects in the data, misspecification can lead to error in the estimation.

Chapter 2 proposes a model in which both the hazard and the frailty distribution are nonparametrically specified as mixtures of B-splines. The B-spline basis is parametrized in such a way that any smooth hazard and frailty density can in principle be estimated, and even non-smooth functions can be approximated very well. Nonparametric modeling of the frailty density substantially decreases the risk of model misspecification, but comes at the cost of additional computational expense. Because of the large number of nuisance parameters involved in fully nonparametric estimation, a maximum likelihood approach is infeasible. Instead,

the proposed model is approached from a Bayesian perspective, and estimation proceeds via Markov Chain Monte Carlo (MCMC). The Bayesian approach has further advantages, notably the ability to estimate any posterior quantity of interest, including the posterior distributions of each of the individual frailties, and posterior hazard quantiles that incorporate uncertainty from all estimated parameters.

Chapter 2 contains a detailed presentation of the model specification, and discusses its relationship to existing frequentist and Bayesian methods for the analysis of clustered survival data. Thorough simulation results expose the advantages and pitfalls of the method. The flexibility of the approach is illustrated using data on rehospitalization due to congestive heart failure, as well as using the diabetic retinopathy data discussed above.

The associated `R` package `splinesurv` implements the methodology and provides a useful tool for statisticians. The package provides sensible default settings that allow it to act as a plug-in replacement for `coxph`, and includes many helpful tools to effectively explore the posterior distributions of commonly interesting quantities. At the time of this writing, the package and documentation are available for download at `http://splinesurv.r-forge.r-project.org`.

1.2 Recurrent event data

In the preceding sections, the event of interest has been implicitly treated as terminal; that is, it was assumed that a subject would be removed from the sample upon experiencing an event. This need not be the case, however.

Processes that generate more than one event per subject are common in applications. For example, Chapter 3 considers a skin cancer study in which multiple lesions may appear on each subject during the course of a study. Other examples of recurrent events include studies of credit defaults, vehicular insurance claims, and assembly line breakdowns.

In the analysis of recurrent event data, the primary objective is to quantify the effect of subject-level covariate information on the risk of event recurrence. As with clustered data, accurate inference must take the structure of the process into account, particularly the dependence between events experienced by a single subject, and the temporal nature of recurrent events. If subjects are additionally clustered, then within-cluster dependence must be considered as well.

An analysis of the skin cancer data should thus estimate the effect of risk factors on the hazard of receiving a lesion, while incorporating possible dependence between the lesions on each subject, and between subjects in the same cluster, while also considering the fact that higher-risk subjects can suffer more lesions during the fixed time-period of the study, and that experiencing some lesions may affect the subsequent appearance of others.

Recurrent event data can be placed into the survival analysis framework discussed above. Cox (1972b) extends the proportional hazards framework to treat recurrent event data as a modulated renewal process, in which the interevent (gap) times are treated as independent conditional on the covariates, which can capture time-varying information such as a count of previous events. Prentice et al. (1981) and Andersen and Gill (1982) extend this approach to allow episode-dependent baseline hazards, time-dependent covariates and strata.

As in the case of clustered data, it is reasonable to suppose that the interevent times experienced by a single subject may have dependence beyond that captured by the measured covariates. Frailty models following the approaches for clustered data can be naturally extended to the recurrent event gap time setting (Aalen and Husebye, 1991; Clayton, 1994; Therneau and Grambsch, 2000). Under certain conditions on the calendar-time dependence structure of the gap times, and with appropriate assumptions on the censoring process, methodology for clustered data can be applied to recurrent event gap time data without modification, although asymptotic justifications differ. A thorough review of existing methodology for recurrent events is provided in Cook and Lawless (2007).

With the inclusion of a frailty, it becomes possible to analyze bivariate or multivariate recurrent event processes, in which subjects experience multiple events of different types. Skin cancer lesions, for example, can be broadly classified into basal cell and squamous cell carcinomas. Risk factors may have different effects on each type of lesion, and the processes for the two types of lesions may depend on each other directly as well as through a frailty. Bivariate models in which both the hazard and frailty distribution are parametrically specified are considered in Abu-Libdeh et al. (1990) and Cook et al. (1999).

Frailty models for recurrent event data face challenges similar to those for clustered data: First, parametric models for the hazard and frailty distribution may not adequately capture the event risk and dependence structure, and are subject to model misspecification risk. Second, the attractive theoretical properties of gamma frailty methods cannot be naturally extended to the common case of clustered recurrent event data, in which subjects experiencing recurrent events are additionally grouped into clusters, thus adding a higher-order layer of dependence.

Lastly, parametric frailty models can require computationally expensive numerical integrations when applied to bivariate or multivariate recurrent event data.

Although a nonparametric approach similar to that of Chapter 2 might also be effective for recurrent event data, this thesis presents a very different method in Chapter 3. Rather than attempting to estimate the frailty distribution nonparametrically, the method avoids specification of the frailty distribution altogether, instead requiring only its first two moments. In this, it follows the approaches of Xue and Brookmeyer (1996) and especially Ma (1999), neither of which require the frailty distribution to be known.

The method proposed in Chapter 3 is an extension of the work of Ma (1999) and Ma et al. (2003) to the setting of bivariate clustered recurrent event data. That is, in addition to accommodating the aforementioned clustered recurrent event case, it allows the events of interest to be of two distinct types, and captures dependence within clusters, within subjects, and between event types by a hierarchical frailty structure that need only be specified through its first two moments. In addition, it incorporates a discretization scheme that reduces the computational effort required for model-fitting, while increasing the stability of the algorithm.

The aforementioned skin cancer data, gathered during the course of the Nutritional Prevention of Cancer study, is used to illustrate the use of the methodology. A second data set, of pulmonary exacerbations experienced by cystic fibrosis patients, serves to demonstrate the applicability of the method to episodic data. All estimation is conducted by means of the associated `blupsurv` package, which at the time of this writing is hosted at `http://blupsurv.r-forge.r-project.org`.

BIBLIOGRAPHY

Aalen, O. and Husebye, E. (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine* **10,** 1227–1240.

Abu-Libdeh, H., Turnbull, B. W., and Clark, L. C. (1990). Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial. *Biometrics* **46,** 1017–1034.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes.* Springer Series in Statistics. Springer-Verlag, New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10,** 1100–1120.

Clayton, D. (1994). Some approaches to the analysis of recurrent event data. *Statistical Methods in Medical Research* **3,** 244–262.

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society. Series A* **148,** 82–117.

Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Event Data.* Springer-Verlag, New York.

Cook, R. J., Ng, E., Mukherjee, J., and Vaughn, D. (1999). Two-state mixed renewal processes for chronic disease. *Statistics in Medicine* **18,** 175–188.

Cox, D. R. (1972a). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34,** 187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.

Cox, D. R. (1972b). The statistical analysis of dependencies in point processes. In Lewis, P., editor, *Stochastic Point Processes*, pages 55–66. Wiley, New York.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73,** 387–396.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. Wiley.

Ma, R. (1999). *An Orthodox BLUP approach to Generalized Linear Models*. PhD thesis, University of British Columbia.

Ma, R., Krewski, D., and Burnett, R. T. (2003). Random effects cox models: A poisson modelling approach. *Biometrika* **90,** 157–169.

McGilchrist, C. A. and Aisbett, C. W. (1991). Regression with frailty in survival analysis. *Biometrics* **47,** 461–466.

Nielsen, G. G., Sørensen, T. I. A., Gill, R. D., and Andersen, P. K. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19,** 25–43.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics* **26,** 183–214.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68,** 373–379.

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

Shih, J. H. and Louis, T. A. (1995). Assessing gamma frailty models for clustered failure time data. *Lifetime Data Analysis* **1,** 205–220.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.

Xue, X. and Brookmeyer, R. (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis* **2,** 277–289.

CHAPTER 2

B-SPLINE FRAILTY MODELS FOR CLUSTERED SURVIVAL DATA

Joint work with David Ruppert and Robert L. Strawderman[1]

The effects of clustering in survival data are commonly addressed by means of frailty models, a generalization of the Cox proportional hazards model (Cox, 1972) in which an unobserved random effect is shared by all members of a cluster (Clayton and Cuzick, 1985), inducing dependence. Frailty models require joint estimation of the regression coefficients, baseline hazard and variance components, because the proportional hazards property required by the partial likelihood method of Cox (1975) does not hold. Under the assumption that frailties follow a gamma or other parametrized distribution, frailty models can be estimated by either choosing a parametric form for the baseline hazard and obtaining a maximum likelihood estimate of all parameters (e.g. Andersen et al., 1993), or by fitting a nonparametric Breslow hazard estimate using an expectation-maximization–type (EM) algorithm (e.g. Murphy, 1995; Parner, 1998; Li et al., 2003).

Extending and building upon the frequentist maximum-likelihood methods, Bayesian approaches to survival analysis have also emerged. Sinha and Dey (1998) propose modeling the baseline hazard by a discrete Lévy process, and Aslanidou et al. (1998) extend this method to a gamma frailty model with hierarchical priors on the gamma dispersion. Parameter estimates can be obtained by Markov Chain Monte Carlo (MCMC), since all conditional posterior distributions can be computed. Related methods include Muliere and Walker (1997); Kim and Lee (2003)

[1]For submission to *Journal of Computational & Graphical Statistics*. Thanks to Mark Cowen and Lakshmi Halasyamani for use of the congestive heart failure data.

and Kottas (2006), which allow different forms of censoring and baseline hazard processes. Such Bayesian approaches, while generally more computationally expensive than their frequentist counterparts, are often more able to accommodate unusual data features, and contain a wealth of information in the form of joint posterior distributions for all parameters of interest. Under the Bayesian paradigm, it is easy to compute any desired posterior quantities from the joint posterior.

All aforementioned traditional and Bayesian methods suffer from the need to specify the distribution of the frailties a priori, and are therefore subject to model error for clustered data problems. Nonparametric Bayesian frailty models have typically avoided modeling the density of the frailties, instead modeling the frailty directly under a Dirichlet process prior (Müller and Quintana, 2004; Pennell and Dunson, 2006), and have retained nonsmooth stepfunction formulations for the baseline hazard that recall the Breslow estimator. The Dirichlet process prior results in discrete posterior frailty distributions, but a frailty with a density may be preferable in many circumstances. Furthermore, simulation results by Barker and Henderson (2005) suggest that Breslow-type baseline hazards may lead to systematic underestimation of the frailty variance and regression parameters, because they depend only on the rank ordering of event times, rather than the actual times. It is therefore desirable to use a method that allows flexible frailty distributions, and nonparametric baseline hazard estimates that incorporate the event times.

In his Ph.D. thesis, Komárek (2006) presents a Bayesian nonparametric approach to clustered survival, based on the accelerated failure time model formulation of Pan (2001) and the random effect density estimation methodology of Ghidey et al. (2004). Both the random effects density and the event time distribution are modeled as smooth mixtures of G-splines and estimated using an MCMC Gibbs

sampler. The resulting error and frailty distributions are smooth, and allow easy visualization and interpretation.

In this chapter, we propose a related model that remains within the proportional hazards framework while implementing the desirable features of the formulation in Komárek and Lesaffre (2006). We model the baseline hazard as a penalized mixture of B-splines, and the frailty density as a penalized mixture of normalized B-splines. Our model formulation and the deconvolution approach bear some resemblance to those of Staudenmayer et al. (2008) and Ruppert et al. (2007), and estimation is similarly carried out using Markov Chain Monte Carlo and a Gibbs sampler for the various model components. The resulting posterior estimates of the baseline hazard and frailty density are smooth and accurate, and can correctly identify unusual frailty densities and baseline hazard forms given sufficiently large samples.

We allow for additional flexibility by proposing two natural extensions: first, we allow the inclusion of a parametric component that may incorporate prior knowledge about the form of the frailty density or baseline hazard. Second, we allow the number and position of knots for the B-spline bases to be chosen adaptively, by a reversible jump MCMC procedure similar to that of Denison et al. (1998) and Biller (2000). Such free-knot spline methods are popular in Bayesian curve-fitting and nonlinear regression (see Smith and Kohn (1996); DiMatteo et al. (2001); Lindstrom (2002) for related work), but have to our knowledge not been adapted to the survival setting. These two extensions respectively improve the performance of the method in situations when the hazard or frailty density can be well-modeled by standard parametric forms, or are distinctly non-smooth.

This chapter is organized as follows: In Section 2.1, we introduce the necessary

notation, present the basic model structure and the aforementioned extensions, and compute the log-posterior parameter density. We give details on the estimation and computer implementation in Section 2.2, including the various conditional posteriors used by each step of the Gibbs sampler, as well as the reversible-jump adaptive knot selection procedure. Section 2.3 contains illustrative simulation results, and discusses the relative merits of our approach and existing ones. In Section 2.4, we apply our methodology to data from studies of congestive heart failure and diabetic retinopathy, and demonstrate the flexibility of the approach. We conclude with a brief discussion in Section 2.5. The appendices contain examples of possible parametric forms and penalties, and the associated computational detail.

2.1 Model structure

We first propose the basic model in Section 2.1.1, which consists of B-spline mixture formulations for the hazard and frailty density curves, with optional penalties to encourage smoothness. In Section 2.1.2 we propose extending the model, by specifying the curves as convex combinations of the aforementioned B-splines and of a parametric component, as a way of incorporating prior knowledge about the hazard or frailty. The extension to adaptive selection of the number and placement of the B-spline knots is presented in Section 2.1.3. Each of the baseline hazard and frailty density curves can then be specified as some combination of a basic model and its optional extensions, each of which can accommodate a variety of priors and options. The result is a flexible family of models that allows prior knowledge about the form and smoothness of either curve to be incorporated into the model fit to the desired extent.

2.1.1 Basic model structure

Suppose that the observed data consists of outcome and covariate information on m independent clusters of size $J_i \geq 1$. We assume that each subject (i, j), $i = 1 \ldots m$, $j = 1 \ldots J_i$ experiences an event at time X_{ij}, and may have available p-dimensional fixed covariates $\boldsymbol{Z}_{ij}$.

Correlations between subjects within the same cluster are captured by a set of m cluster-level frailties U_i, $i = 1 \ldots m$, which are positive, and independent and identically distributed with a density f, whose mean is 1. When the variance of f is zero, all frailties are fixed at 1, and the problem reduces to the setting of unclustered survival analysis.

Each subject's event time may be censored at a time C_{ij}, which is assumed to be noninformative in the sense of Nielsen et al. (1992). Denote the followup time as $T_{ij} = X_{ij} \wedge C_{ij}$ and the status indicator as $\delta_{ij} = I(X_{ij} < C_{ij})$.

Denote the baseline hazard by $\lambda_0(t)$ with corresponding cumulative hazard $\Lambda_0(t) = \int_0^t \lambda(s)ds$. Following the proportional hazards framework, the hazard and cumulative hazard for T_{ij} conditional on the frailties and regression coefficients are

$$\lambda_{ij}(t|\boldsymbol{U}, \boldsymbol{Z}) = U_i \lambda_0(t) e^{\boldsymbol{Z}_{ij}^T \boldsymbol{\beta}} \quad \text{and} \quad \Lambda_{ij}(t|\boldsymbol{U}, \boldsymbol{Z}) = U_i \Lambda_0(t) e^{\boldsymbol{Z}_{ij}^T \boldsymbol{\beta}} \,,$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of regression coefficients. The relevant likelihood conditional on the frailties (see Andersen et al., 1993) can then be written as

$$\mathcal{L}(\boldsymbol{\beta}, \lambda|\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{U}, \boldsymbol{Z}) = \prod_{i=1}^{m} \prod_{j=1}^{J_i} \frac{\left(U_i \lambda_0(T_{ij}) e^{\boldsymbol{Z}_{ij}^T \boldsymbol{\beta}}\right)^{\delta_{ij}}}{\exp\left(U_i \Lambda_0(T_{ij}) e^{\boldsymbol{Z}_{ij}^T \boldsymbol{\beta}}\right)} \,. \tag{2.1}$$

Specifically, we initially model the baseline hazard as a non-negative linear combination of K_λ B-spline basis functions $B_{\lambda k}(x)$ of order Q_λ, defined on $N_\lambda =$

$K_\lambda - Q_\lambda$ interior knots $\boldsymbol{\xi}_\lambda$ distributed over the range of the event times. The splines are indexed by parameters $\boldsymbol{\theta}_\lambda$, with the weight of each spline basis function given by $w_{\lambda k} = e^{\theta_{\lambda k}}$. That is, the hazard and cumulative hazard can be written as

$$\lambda_0(t|\boldsymbol{\theta}_\lambda) = \sum_{k=1}^{K_\lambda} B_{\lambda k}(t) w_{\lambda k} \,, \qquad \Lambda_0(t|\boldsymbol{\theta}_\lambda) = \sum_{k=1}^{K_\lambda} w_{\lambda k} \int_0^t B_{\lambda k}(s) ds \,. \tag{2.2}$$

Similarly, we initially model the frailty density f as a convex combination of normalized B-spline basis functions $\tilde{B}_{uk}(x)$ of order Q_u, defined on knots $\boldsymbol{\xi}_u$ over a sufficiently large range. The splines are indexed by parameters $\boldsymbol{\theta}_u$, and both the B-splines and weights are normalized to ensure that the density integrates to 1. That is, we assume the frailties are independent and identically distributed as $U_i|\boldsymbol{\theta}_u \sim f(x|\boldsymbol{\theta}_u)$, with

$$f(x|\boldsymbol{\theta}_u) = \sum_{k=1}^{K_u} \tilde{B}_{uk}(x) w_{uk} \,, \tag{2.3}$$

where

$$w_{uk} = \frac{\exp(\theta_{uk})}{\sum_{\ell=1}^{K_u} \exp(\theta_{u\ell})} \,, \qquad \text{and} \qquad \tilde{B}_{uk}(x) = B_{uk}(x) \cdot \left(\int_{-\infty}^{\infty} B_{uk}(s) ds \right)^{-1} \,.$$

As in Staudenmayer et al. (2008), we place a multivariate Normal prior on the regression parameters $\boldsymbol{\beta}$. Priors on the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_\lambda, \boldsymbol{\theta}_u)$ have a Gaussian structure, but may incorporate a penalty to induce smoothness in the B-spline coefficients and avoid overfitting. Denoting the penalty functions as $p_\lambda(\boldsymbol{\theta}_\lambda)$ and $p_u(\boldsymbol{\theta}_u)$, the priors can then be written as:

$$\pi(\boldsymbol{\beta}|\sigma_\beta^2) = (2\pi\sigma_\beta^2)^{-\frac{p}{2}} \exp\left(-\frac{1}{2\sigma_\beta^2} \boldsymbol{\beta}^T \boldsymbol{\beta} \right) \,, \tag{2.4}$$

$$\pi(\boldsymbol{\theta}_\lambda|\sigma_\lambda^2) \propto (2\pi\sigma_\lambda^2)^{-\frac{K_\lambda}{2}} \exp\left(-\frac{1}{2\sigma_\lambda^2} p_\lambda(\boldsymbol{\theta}_\lambda) \right) \,, \tag{2.5}$$

$$\pi(\boldsymbol{\theta}_u|\sigma_u^2) \propto (2\pi\sigma_u^2)^{-\frac{K_u}{2}} \exp\left(-\frac{1}{2\sigma_u^2} p_u(\boldsymbol{\theta}_u) \right) \,. \tag{2.6}$$

Depending on the choice of penalty functions p_u, p_λ, the priors for $\boldsymbol{\theta}_\lambda$ and $\boldsymbol{\theta}_u$ may be improper. They may be chosen to follow a simple Gaussian form similar to eq. (2.4), or to penalize second differences in the parameters $\boldsymbol{\theta}_u, \boldsymbol{\theta}_\lambda$, or the integrated squared second derivative of the spline, or other smoothness criteria. Examples of such penalty functions are presented in appendix 2.A.

Lastly, assume inverse-gamma priors for the error variance parameters $\sigma_\beta^2, \sigma_\lambda^2, \sigma_u^2$, with parameters $\boldsymbol{\alpha}_\beta = (\alpha_{\beta 1}, \alpha_{\beta 2})$ (and analogously for the others), so that

$$\pi(\sigma_\beta^2 | \boldsymbol{\alpha}_\beta) \propto (\sigma_\beta^2)^{-(\alpha_{\beta 1}+1)} \exp\left(-\frac{\alpha_{\beta 2}}{\sigma_\beta^2}\right) , \tag{2.7}$$

and analogously for σ_λ^2 and σ_u^2. These may be quite diffuse.

Then, the log-posterior (up to a constant) can be written as

$$\begin{aligned}
&\ell(\boldsymbol{U}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\alpha} | \boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Z}) = \\
&\quad \sum_{i,j} \delta_{ij} \left(\ln U_i + \ln \lambda_0(T_{ij}|\boldsymbol{\theta}_\lambda) + \boldsymbol{Z}_{ij}^T \boldsymbol{\beta} \right) - \sum_{i,j} U_i \Lambda_0(T_{ij}|\boldsymbol{\theta}_\lambda) e^{\boldsymbol{Z}_{ij}^T \boldsymbol{\beta}} + \sum_i \ln f(U_i|\boldsymbol{\theta}_u) \\
&\quad + \left(-\frac{p}{2} \ln \sigma_\beta^2 - \frac{\boldsymbol{\beta}^T \boldsymbol{\beta}}{2\sigma_\beta^2} \right) + \left(-\frac{K_\lambda}{2} \ln \sigma_\lambda^2 - \frac{p_\lambda(\boldsymbol{\theta}_\lambda)}{2\sigma_\lambda^2} \right) + \left(-\frac{K_u}{2} \ln \sigma_u^2 - \frac{p_u(\boldsymbol{\theta}_u)}{2\sigma_u^2} \right) \\
&\quad -(\alpha_{\beta 1}+1) \ln \sigma_\beta^2 - \frac{\alpha_{\beta 2}}{\sigma_\beta^2} - (\alpha_{\lambda 1}+1) \ln \sigma_\lambda^2 - \frac{\alpha_{\lambda 2}}{\sigma_\lambda^2} - (\alpha_{u 1}+1) \ln \sigma_u^2 - \frac{\alpha_{u 2}}{\sigma_u^2}
\end{aligned}$$

In this expression, the terms respectively correspond to the point process likelihood in eq. (2.1) and the frailty density in eq. (2.3), the spline parameter priors eq. (2.4)–(2.6), and eq. (2.7) and its analogues.

2.1.2 Addition of a parametric component

For smaller samples, the dimensionality of the B-spline parametrization of the baseline hazard and frailty density may be a liability. When the data itself contains

relatively little information about the shape of the curve, the Bayesian approach makes it possible to supplement the data with prior knowledge. It is, however, not intuitive to define an informative prior on the spline parameters $\boldsymbol{\theta}$.

We thus propose a modification to the methodology of Section 2.1.1 that extends the aforementioned spline model by including a parametrically specified basis function. For the baseline hazard, the parametric component might take the form of a Weibull or lognormal family, and for the frailty density one might consider a gamma or lognormal component. Specifically, the form of the baseline hazard in eq. (2.2) is replaced by a convex combination of the spline component and a new parametric component,

$$\begin{aligned}\lambda_0(t|\boldsymbol{\theta}_\lambda, \boldsymbol{\eta}_\lambda, \phi_\lambda) &= \phi_\lambda \sum_{k=1}^{K_\lambda} B_{\lambda k}(t) w_{\lambda k} + (1-\phi_\lambda)\lambda_{0p}(t|\boldsymbol{\eta}_\lambda) \\ \Lambda_0(t|\boldsymbol{\theta}_\lambda, \boldsymbol{\eta}_\lambda, \phi_\lambda) &= \phi_\lambda \sum_{k=1}^{K_\lambda} w_{\lambda k} \int_0^t B_{\lambda k}(s)ds + (1-\phi_\lambda)\Lambda_{0p}(t|\boldsymbol{\eta}_\lambda) \ ,\end{aligned} \tag{2.8}$$

and the specification for the frailty density in eq. (2.3) is replaced by

$$f(x|\boldsymbol{\theta}_u, \boldsymbol{\eta}_u, \phi_u) = \phi_u \sum_{k=1}^{K_u} \tilde{B}_{uk}(x) w_{uk} + (1-\phi_u) f_p(x|\boldsymbol{\eta}_u) \ , \tag{2.9}$$

where ϕ_λ, ϕ_u represent the weights of the nonparametric components, $\lambda_{0p}, \Lambda_{0p}$ and f_p are respectively the parametric baseline hazard, cumulative hazard and frailty density components, and $\boldsymbol{\eta}_\lambda, \boldsymbol{\eta}_u$ are the parameters indexing these components.

Intuitively then, the formulation in eq. 2.8 may be viewed as a parametric baseline hazard, with deviations captured by a spline component. The prior on the weight ϕ_λ specifies the degree of confidence in the parametric component, increasing with smaller values of ϕ_λ. A prior favoring the parametric component ensures that in data-poor circumstances, the fit shrinks towards a parametric specification.

This can be accomplished by placing Beta priors on the weights $\boldsymbol{\phi} = (\phi_\lambda, \phi_u)$, with fixed hyperparameters $\boldsymbol{\alpha}_{\phi_\lambda}, \boldsymbol{\alpha}_{\phi_u}$. A Beta$(1,1)$ (uniform) prior is inadequate, because, the weights $\boldsymbol{\phi}$ may not be identifiable when the spline and parametric components have similar forms. However, a nonuniform prior such as Beta$(1,2)$ (triangular) ensures that the weight of the nonparametric component shrinks if it does not capture information beyond that captured by the parametric portion. Alternatively, a Beta$(2,1)$ prior can place additional weight on the nonparametric component if an unusual structure is suspected, while still allowing the possibility of very low frailty variance. Fixing the weights at 1 reduces the model to that of Section 2.1.1, whereas fixing the weights at 0 leads to a purely parametric Bayesian survival model.

Priors for the parametric terms $\boldsymbol{\eta}_\lambda$ and $\boldsymbol{\eta}_u$ necessarily depend on the desired parametric form. Denote the priors by $\pi_\lambda(\boldsymbol{\eta}_\lambda|\boldsymbol{\sigma}^2_{\eta_\lambda})$ and $\pi_u(\boldsymbol{\eta}_u|\boldsymbol{\sigma}^2_{\eta_u})$, where $\boldsymbol{\sigma}^2_{\eta_\lambda}, \boldsymbol{\sigma}^2_{\eta_u}$ themselves may have priors depending on hyperparameters $\boldsymbol{\alpha}_{\boldsymbol{\eta}} = (\boldsymbol{\alpha}_{\eta_\lambda}, \boldsymbol{\alpha}_{\eta_u})$. In practice, we have found it effective to parametrize the distributions λ_{0p}, f_p in a way that permits Gaussian priors. A few reasonable choices for common parametric forms are presented in appendix 2.B.

The posterior loglikelihood is then given by:

$$\ell(\boldsymbol{U}, \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{\alpha}|\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Z}) = \ell(\boldsymbol{U}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\alpha}|\boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Z}) \tag{2.10}$$

$$+ \log\left[\phi_\lambda^{(\alpha_{\phi_\lambda 1}-1)}(1-\phi_\lambda)^{(\alpha_{\phi_\lambda 2}-1)}\right] + \log\left[\phi_u^{(\alpha_{\phi_u 1}-1)}(1-\phi_u)^{(\alpha_{\phi_u 2}-1)}\right] \tag{2.11}$$

$$+ \log \pi_\lambda(\boldsymbol{\eta}_\lambda|\boldsymbol{\sigma}^2_{\eta_\lambda}) + \log \pi_u(\boldsymbol{\eta}_u|\boldsymbol{\sigma}^2_{\eta_u}) + \log \pi_\eta(\boldsymbol{\sigma}^2_{\eta_\lambda}, \boldsymbol{\sigma}^2_{\eta_u}|\boldsymbol{\alpha}_\eta) \tag{2.12}$$

where $\ell(\boldsymbol{U}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\alpha}|\boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Z})$ is analogous to eq. (2.8)–(2.8), with eqs. (2.8) and (2.9) substituted for the hazard and frailty curves. The terms in (2.11) contain the Beta priors on the weights, and terms (2.12) contain the priors for the parametric components.

2.1.3 Adaptive knot selection

Thus far, little has been said about the choice of the number of spline knots N_λ, N_u and their positions $\boldsymbol{\xi}_\lambda, \boldsymbol{\xi}_u$. The number and placement of knots has a profound effect on the smoothness of the estimated hazard curve or frailty density: including only few widely-spaced knots generally leads to very smooth curves, whereas for multiple knots in close proximity, smoothness has to be enforced by a penalty so as to avoid the risk of overfitting.

In the work of Staudenmayer et al. (2008) and Komárek and Lesaffre (2006), as in the development of the preceding sections, the number of knots is fixed, and their positions are distributed evenly over the range of the data. If the underlying curve is smooth, this approach yields excellent results in conjunction with smoothing penalties in the priors of eq. (2.5) and (2.6), provided that the knots and smoothing penalties are well-chosen. In the survival setting however, such specification can be particularly challenging, since the hazard is observable only through the event times, and its smoothness may be difficult to judge. Furthermore, most penalized smoothing acts globally over the range of the data, but there is often no a priori reason to suspect that all regions of the hazard or frailty curve need to be similarly smooth.

An advantage of the Bayesian approach is that the number and positions of knots may be treated as additional parameters to be estimated. In that case, the likelihood in eq. (2.8)–(2.8) should be treated as the conditional likelihood given the number of knots $\boldsymbol{N} = (N_\lambda, N_u)$ and their positions $\boldsymbol{\xi} = (\boldsymbol{\xi}_\lambda, \boldsymbol{\xi}_u)$.

Biller (2000) introduced an approach for automatic knot selection for generalized linear models using natural cubic splines, in which the number of knots and

spline weights were chosen by reversible-jump MCMC methods (Green, 1995), similar to the curve-fitting procedure of Denison et al. (1998). The procedure makes it possible to sample from the posterior of the model set consisting of different numbers and placements of knots, and corresponding different dimensionality of the spline parameters.

A similar approach can be used in this case to extend the spline formulation presented in Section 2.1.1, with the added complexities of the survival model specification and frailty deconvolution. In addition to the Gibbs sampling steps required to sample from the posterior of the parameters introduced in the preceding sections, the adaptive knot selection procedure consists of three possible moves that affect the hazard and frailty density curves by changing their B-spline basis: the addition of a knot, the removal of a knot, and the change in position of a knot.

When selecting the knots adaptively, the smoothness of the curve can be dictated simply by the number of knots and their positions, with no need for an additional smoothing penalty. In this case, the prior on the number of knots $\boldsymbol{N}$ plays a key role in specifying the smoothness. Denison et al. (1998), suggests placing Poisson priors on the number of knots, and simulation experiments conducted by Biller (2000) indicate that this gives good results in the context of nonlinear regression. The Poisson prior is strongly informative, and allows great control over the smoothness of the resulting curve, at the risk of overfitting. In contrast, less informative priors such as a Geometric or Negative Binomial can be used to penalize large numbers of knots and give preference to smoother curves. We consider these and other priors as well.

Given the number of knots $\boldsymbol{N}$, they may then take positions uniformly on a much larger set of M_λ, M_u predetermined candidate knot positions $\boldsymbol{\xi}_\lambda^c =$

$\{\xi^c_{\lambda 1}, \ldots, \xi^c_{\lambda M_\lambda}\}$, $\boldsymbol{\xi}^c_u = \{\xi^c_{u1}, \ldots, \xi^c_{uM_u}\}$. The choice of candidate knots represents a prior on the knot positions: for instance, for the baseline hazard, candidate knots might be selected as quantiles of the observed event times, to make data-rich regions more likely to contain knots.

The joint log-posterior likelihood up to a constant may then be written as:

$$\ell(\boldsymbol{U}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\xi}, \boldsymbol{N} | \boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Z}) = \ell(\boldsymbol{U}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\alpha} | \boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Z}, \boldsymbol{\xi}, \boldsymbol{N}) + \log \pi(\boldsymbol{\xi} | \boldsymbol{N}) + \log \pi(\boldsymbol{N}) \ , \tag{2.13}$$

where $\ell(\boldsymbol{U}, \boldsymbol{\theta}, \boldsymbol{\sigma}, \boldsymbol{\alpha} | \boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{Z}, \boldsymbol{\xi}, \boldsymbol{N})$ is as in eq. (2.8)–(2.8), and the remaining terms represent the aforementioned priors on knot position and parameter dimension respectively. Parametric components may be included as well, in the same way as discussed in Section 2.1.2.

We will discuss the details of this reversible-jump MCMC step in Section 2.2.4.

2.2 Estimation procedure

The estimation procedure is a multi-step algorithm consisting of three types of steps: Initialization, parameter updates via Gibbs sampling and Metropolis-Hastings MCMC, and, if desired, adaptive knot selection via reversible-jump MCMC. After initialization, posterior samples of all parameter can be drawn by ordinary and reversible-jump MCMC steps, repeated as long as needed to ensure convergence of the chain and a sufficient number samples from the posterior. Figure 2.1 shows the structure of the algorithm. We propose additional notation in Section 2.2.1, and discuss initialization in Section 2.2.2. Each of the Metropolis-Hastings parameter update steps is presented in Section 2.2.3, and the reversible-jump method used for knot selection is discussed in Section 2.2.4.

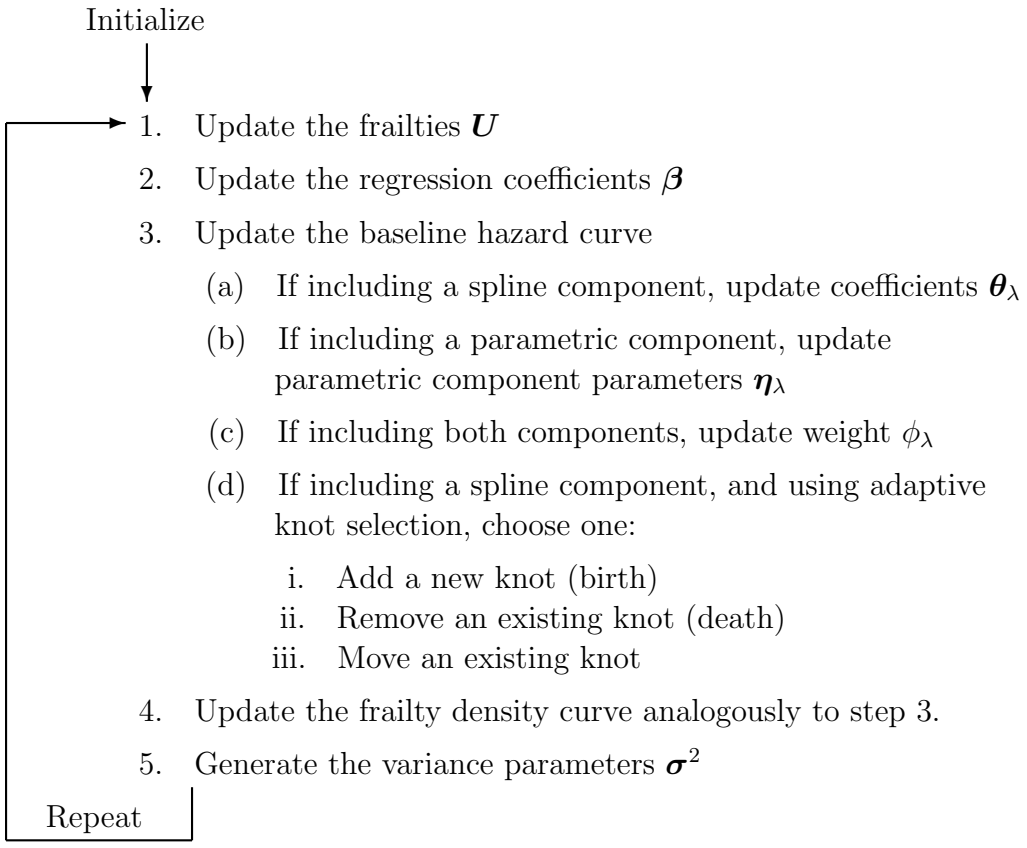

Figure 2.1: Structure of the estimation procedure.

2.2.1 Additional notation

In the following, denote by $\boldsymbol{\beta}^{(h)}$ the estimated regression parameters at iteration h, with initial values corresponding to $h = 0$, and analogously for the frailties and other estimated parameters $\boldsymbol{U}^{(h)}, \boldsymbol{\theta}_\lambda{}^{(h)}, \boldsymbol{\theta}_u{}^{(h)}, \boldsymbol{\eta}_\lambda^{(h)}, \boldsymbol{\eta}_u^{(h)}, \boldsymbol{\phi}^{(h)}, \boldsymbol{\sigma}^{2(h)}$. Also, denote the spline parameter weight vectors $\boldsymbol{w}_\lambda^{(h)} = e^{\boldsymbol{\theta}_\lambda^{(h)}}$ and $\boldsymbol{w}_u^{(h)} = e^{\boldsymbol{\theta}_u^{(h)}} / \mathbf{1}_m^T e^{\boldsymbol{\theta}_u^{(h)}}$.

If adaptive knot selection is used, the number and placement of knots may also vary with each iteration. Denote the number of knots $N_\lambda^{(h)}, N_u^{(h)}$ and corresponding spline parameter dimensions $K_\lambda^{(h)}, K_u^{(h)}$. Knots are located at $\boldsymbol{\xi}_\lambda^{(h)}, \boldsymbol{\xi}_u^{(h)}$, and correspond to B-spline bases $B_{\lambda k}^{(h)}(t)$ and $\tilde{B}_{uk}^{(h)}(x)$ for the hazard and frailty respectively.

Without adaptive knot selection, $B_{\lambda k}^{(h)}(t) = B_{\lambda k}^{(0)}(t)$, and analogous for the frailty splines.

Further, define basis function vectors $\tilde{\boldsymbol{B}}_u^{(h)}(x) = \left(\tilde{B}_{u1}^{(h)}(x), \ldots, \tilde{B}_{uK_u^{(h)}}^{(h)}(x)\right)$ and analogously for $\boldsymbol{B}_\lambda^{(h)}(t)$. Additionally, denote

$$C_{\lambda k}^{(h)}(t) = \int_0^t B_{\lambda k}^{(h)}(s)ds\,, \quad \text{and} \quad E_{uk}^{(h)} = 1 - \int_{-\infty}^{\infty} x\tilde{B}_{uk}^{(h)}(x)dx\,, \tag{2.14}$$

with corresponding vectors $\boldsymbol{C}_\lambda^{(h)}(t)$ and $\boldsymbol{E}_u^{(h)}$. The vector $\boldsymbol{E}_u^{(h)}$ is used in the course of parameter estimation to ensure that the frailty mean is 1, for identifiability. We give simple recursive formulas for these integrals in appendix 2.C.

To make possible full matrix notation, we define the following vectors and matrices:

$$\tilde{\boldsymbol{U}}^{(h)} = \left[U_{11}^{(h)} \ldots U_{mJ_m}^{(h)}\right]^T \qquad \boldsymbol{\delta} = [\delta_{11} \ldots \delta_{mJ_m}]^T \qquad \boldsymbol{Z} = [\boldsymbol{Z}_{11} \ldots \boldsymbol{Z}_{mJ_m}]^T$$

$$\tilde{\boldsymbol{B}}_u^{(h)} = \begin{bmatrix} \tilde{\boldsymbol{B}}_u(U_1^{(h)}) \\ \vdots \\ \tilde{\boldsymbol{B}}_u(U_m^{(h)}) \end{bmatrix} \quad \boldsymbol{B}_\lambda^{(h)} = \begin{bmatrix} \boldsymbol{B}_\lambda^{(h)}(T_{11}) \\ \vdots \\ \boldsymbol{B}_\lambda^{(h)}(T_{mJ_m}) \end{bmatrix} \quad \boldsymbol{C}_\lambda^{(h)} = \begin{bmatrix} \boldsymbol{C}_\lambda^{(h)}(T_{11}) \\ \vdots \\ \boldsymbol{C}_\lambda^{(h)}(T_{mJ_m}) \end{bmatrix}$$

$$\boldsymbol{\lambda}_p^{(h)} = \begin{bmatrix} \lambda_p(T_{11}|\boldsymbol{\eta}_\lambda^{(h)}) \\ \vdots \\ \lambda_p(T_{mJ_m}|\boldsymbol{\eta}_\lambda^{(h)}) \end{bmatrix} \quad \boldsymbol{\Lambda}_p^{(h)} = \begin{bmatrix} \Lambda_p(T_{11}|\boldsymbol{\eta}_\lambda^{(h)}) \\ \vdots \\ \Lambda_p(T_{mJ_m}|\boldsymbol{\eta}_\lambda^{(h)}) \end{bmatrix} \quad \boldsymbol{f}_p^{(h)} = \begin{bmatrix} f_p(U_1|\boldsymbol{\eta}_u^{(h)}) \\ \vdots \\ f_p(U_m|\boldsymbol{\eta}_u^{(h)}) \end{bmatrix}.$$

Note that in $\tilde{\boldsymbol{U}}^{(h)}$, we use $U_{ij}^{(h)} = U_i^{(h)}$, giving a vector of length $\sum_i J_i$ (in constrast, $\boldsymbol{U}^{(h)}$ is of length m). The matrix $\boldsymbol{Z}$ has dimensions $\sum_i J_i \times p$, $\tilde{\boldsymbol{B}}_u^{(h)}$ has dimension $m \times K_u^{(h)}$, and matrices $\boldsymbol{B}_\lambda^{(h)}, \boldsymbol{C}_\lambda^{(h)}$ are $\sum J_i \times K_\lambda^{(h)}$. We will also denote by $D(\boldsymbol{x})$ a diagonal matrix with $\boldsymbol{x}$ on the diagonal.

This allows us to construct at each iteration vectors of the estimated baseline hazard at each observed event time, and of the frailty density at each estimated

frailty, corresponding to evaluations of equations (2.8) and (2.9):

$$\begin{aligned}
\boldsymbol{\lambda}_0^{(h)} &= \phi_\lambda^{(h)} \boldsymbol{B}_\lambda^{(h)} \boldsymbol{w}_\lambda^{(h)} + (1 - \phi_\lambda^{(h)}) \boldsymbol{\lambda}_{0p}^{(h)} \\
\boldsymbol{\Lambda}_0^{(h)} &= \phi_\lambda^{(h)} \boldsymbol{C}_\lambda^{(h)} \boldsymbol{w}_\lambda^{(h)} + (1 - \phi_\lambda^{(h)}) \boldsymbol{\Lambda}_{0p}^{(h)} \\
\boldsymbol{f}^{(h)} &= \phi_u^{(h)} \tilde{\boldsymbol{B}}_u^{(h)} \boldsymbol{w}_u^{(h)} + (1 - \phi_u^{(h)}) \boldsymbol{f}_p^{(h)} \ ,
\end{aligned}$$

The dependence of these estimates on other parameters will be used implicitly in the construction of conditional likelihoods for the Metropolis-Hastings steps.

Lastly, define tuning parameters $\gamma_\nu, \gamma_\lambda, \gamma_\beta, \gamma_u$ chosen to make acceptance probabilities in the Metropolis-Hastings algorithm close to 25%. These may be set manually or can be selected by an adaptive procedure during the burn-in phase of the chain. We present a simplistic but effective adaptive procedure to choose MCMC tuning parameters during the burn-in phase in Section 2.2.3.9.

2.2.2 Obtaining initial values

Even though the chain can in principle be initialized at any value, we find that good starting values hasten the convergence of the chain, and reduce the risk of numerical problems. This section contains initial values that we have found to yield good results. They are in part found by using computationally inexpensive frequentist methods, and by maximizing conditional likelihoods.

For the the B-spline $\boldsymbol{B}_\lambda(t)$ specifying the baseline hazard, with events occurring at times T_{ij}, we set the initial number of interior knots to

$$N_\lambda^{(0)} = \min\left(\frac{\sum_i J_i}{4}, 35\right) \ ,$$

distributed evenly or by quantiles along the range of observed event times. In addition, if the spline order $Q_\lambda > 1$, we define repeated exterior knots located at

the boundaries, to ensure that $\boldsymbol{B}_\lambda^{(h)}(t)$ is supported exactly on the range of event times.

Analogously, for the (unnormalized) B-spline $\boldsymbol{B}_u^{(h)}(x)$ of order Q_u defining the frailty density, we choose a support range $(U_{\min}, U_{\max})$, and distribute $N_u^{(0)} = \min(m/4, 35)$ knots evenly across the range. The range may be chosen a priori, or it may be allowed to depend on the initial estimates of the frailties from a parametric frailty model.

We set the initial values for all parameters as follows:

- Hyperparameters $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_\beta, \boldsymbol{\alpha}_\lambda, \boldsymbol{\alpha}_u)$ are fixed as

$$\alpha_{\beta 1} = \alpha_{\beta 2} = \alpha_{\lambda 1} = \alpha_{\lambda 2} = \alpha_{u1} = \alpha_{u2} = 0.01 \ ,$$

  indicating diffuse priors on the regression and spline parameters. Hyperparameters $\boldsymbol{\alpha}_\eta$ may depend on the parametric form chosen, but for the parameterizations given in Section 2.B, analogous diffuse Gaussian priors are appropriate. The hyperparameters $\boldsymbol{\alpha}_{\phi_\lambda}, \boldsymbol{\alpha}_{\phi_u}$ determine the form of the Beta prior on the weights ϕ_λ, ϕ_u.

- Initial values for the frailty estimates $\boldsymbol{U}^{(0)}$ and for the parameter estimates $\boldsymbol{\beta}^{(0)}$ can be obtained by fitting a frequentist proportional hazards frailty model to the data (e.g. via `coxph` in `R`). The frailty distribution should match the parametric component, if available.

- Parametric baseline hazard parameters $\boldsymbol{\eta}_\lambda$ can be initialized by fitting a parametric proportional hazards model to the data (e.g. via `survreg` in `R`). Initial values for the parametric frailty density parameters $\boldsymbol{\eta}_u$ can be obtained from the frailty model fit previously.

- Initial values for the variance parameters ${\boldsymbol{\sigma}^2}^{(0)}$ are set as

$$\sigma_\beta^{2(0)} = \sigma_\lambda^{2(0)} = \sigma_\lambda^{2(0)} = \sigma_{\eta_\lambda}^{2(0)} = \sigma_{\eta_u}^{2(0)} = 0.1 \; .$$

- Initial values for the spline coefficients ${\boldsymbol{\theta}_\lambda}^{(0)}, {\boldsymbol{\theta}_u}^{(0)}$ can be found by maximizing the conditional log-likelihoods $\ell(\boldsymbol{\theta}_\lambda|\boldsymbol{U}^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\phi}^{(0)}, {\boldsymbol{\sigma}^2}^{(0)})$ and $\ell(\boldsymbol{\theta}_u|\boldsymbol{U}^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\eta}^{(0)}, \boldsymbol{\phi}^{(0)}, {\boldsymbol{\sigma}^2}^{(0)})$ (see Section 2.2.3.3 and Section 2.2.3.4 for the likelihoods. Formulas for the gradients are given in appendix 2.D.

2.2.3 Metropolis-Hastings MCMC steps

Once initial values have been obtained, the algorithm begins running the MCMC sampling loop for a specified number of iterations. The first stage consists of successive Gibbs sampling steps, in which, given iteration h parameter estimates, each set of parameters is updated in turn by Metropolis-Hastings. The following subsections detail the likelihoods for each of these parameters conditional on the remainder, which for brevity are denoted by an ellipsis.

2.2.3.1 Updating the frailties U_i

The loglikelihood for the frailty parameters $\boldsymbol{U}$ conditional on the remaining parameters is given by portions of component (2.8) in the full likelihood, that is:

$$\ell\left(\boldsymbol{U}|\ldots\right) \;=\; \sum_{i=1}^{m}\left[\sum_{j=1}^{J_i}\left(\delta_{ij}\log U_i - U_i\boldsymbol{\Lambda}_0^{(h)}e^{\boldsymbol{Z}_{ij}^T\boldsymbol{\beta}^{(h)}}\right) + \log f_i^{(h)}\right]$$

Clearly, conditional on the remaining parameters, the frailties U_i are independent. We thus generate candidates individually for the $h+1$-st iteration from a Gamma

transition kernel with mean $U_i^{(h)}$ and variance γ_ν, that is,

$$\tilde{U}_i^{(h+1)} \sim \text{Gamma}\left\{\gamma_\nu^{-1}\left(U_i^{(h)}\right)^2, \gamma_\nu\left(U_i^{(h)}\right)^{-1}\right\} ,$$

where γ_ν is a tuning parameter. Each candidate is accepted with probability

$$\rho_u = \min\left\{1, \frac{\mathcal{L}(\tilde{U}_i^{(h+1)}|\ldots)\mathcal{T}(\tilde{U}_i^{(h+1)}, U_i^{(h)})}{\mathcal{L}(U_i^{(h)}|\ldots)\mathcal{T}(U_i^{(h)}, \tilde{U}_i^{(h+1)})}\right\} ,$$

where $\mathcal{L}$ is the posterior likelihood, $\mathcal{T}(x, x')$ is a gamma transition kernel with mean x and variance γ_ν evaluated at x'.

2.2.3.2 Updating regression coefficients β

The loglikelihood for the regression coefficients conditional on the remaining parameters can be written as

$$\ell\left(\boldsymbol{\beta}|\ldots\right) \;=\; \boldsymbol{\delta}^T\boldsymbol{Z}\boldsymbol{\beta} - \tilde{\boldsymbol{U}}^{(h)^T}D(\boldsymbol{\Lambda}_0^{(h)})e^{\boldsymbol{Z}\boldsymbol{\beta}} - \frac{\boldsymbol{\beta}^T\boldsymbol{\beta}}{2\sigma_\beta^2} . \tag{2.15}$$

Candidates for the $h+1$-st iteration can be generated from $N\left(\boldsymbol{\beta}^{(h)}, \gamma_\beta\boldsymbol{\Sigma}_\beta\right)$, where $\boldsymbol{\Sigma}_\beta$ is the inverse Hessian of the likelihood in eq. (2.15) evaluated at the initial values (see eq. (2.24)).

2.2.3.3 Updating baseline hazard spline coefficients θ_λ

The nonparametric estimate of the baseline hazard depends on parameters $\boldsymbol{\theta}_\lambda$ as detailed in eq. (2.8). The loglikelihood for these coefficients conditional on the remaining parameters is given by

$$\ell\left(\boldsymbol{\theta}_\lambda|\ldots\right) = \boldsymbol{\delta}^T\log\boldsymbol{\lambda}_0^{(h)} - \tilde{\boldsymbol{U}}^{(h)^T}D(\boldsymbol{\Lambda}_0^{(h)})e^{\boldsymbol{Z}\boldsymbol{\beta}^{(h)}} - \frac{p_\lambda(\boldsymbol{\theta}_\lambda)}{2\sigma_\lambda^2} \tag{2.16}$$

Candidates for the $h+1$-st iteration can be generated one by one from $N\left(\theta_{\lambda k}^{(h)}, \gamma_\lambda\right)$, where γ_λ is a tuning parameter.

2.2.3.4 *Updating frailty density spline coefficients θ_u*

The loglikelihood for the frailty density spline coefficients $\boldsymbol{\theta}_u$ conditional on the remaining parameters is given by

$$\ell\left(\boldsymbol{\theta}_u|\boldsymbol{U},\boldsymbol{\beta},\boldsymbol{\theta}_\lambda\right) \;=\; \mathbf{1}_m^T \log \boldsymbol{f}^{(h)} - \frac{p_u(\boldsymbol{\theta}_u)}{2\sigma_u^2} \tag{2.17}$$

In generating candidates for the $h+1$-st iteration, it is important for identifiability to ensure that the mean of the frailty density is fixed at 1. This condition can be expressed as the constraint

$$\boldsymbol{E}_u^{(h)} e^{\boldsymbol{\theta}_u^{(h)}} = 0 \; .$$

where $\boldsymbol{E}_u^{(h)}$ is defined in eq. (2.14). We therefore generate candidates in pairs, in such a way that the constraint is always satisfied: for each k, we first generate a candidate $\tilde{\theta}_{uk}^{(h+1)} \sim N\left(\boldsymbol{\theta}_{uk}^{(h)}, \gamma_u\right)$, where γ_u is a tuning parameter. We then choose a random index m among the remainder, and adjust the candidate $\tilde{\theta}_{um}^{(h+1)}$ as

$$\tilde{\theta}_{um}^{(h+1)} = \log\left\{ \frac{1}{E_{um}^{(h)}} \left(E_{uk}^{(h)} (e^{\theta_{uk}^{(h)}} - e^{\tilde{\theta}_{uk}^{(h+1)}}) + E_{um}^{(h)} e^{\theta_{um}^{(h)}} \right) \right\} \; ,$$

which ensures that the pair $(\theta_{uk}^{(h+1)}, \theta_{um}^{(h+1)})$ continues to satisfy the constraint. Since this corresponds to a symmetric transition kernel, each pair is accepted with a standard Metropolis probability.

2.2.3.5 *Updating baseline hazard parametric component parameters η_λ*

The parametric component λ_{0p} of the baseline hazard λ_0 depends on parameters $\boldsymbol{\eta}_\lambda$. The loglikelihood for these parameters is given by

$$\ell\left(\boldsymbol{\eta}_\lambda|\ldots\right) = \boldsymbol{\delta}^T \log \boldsymbol{\lambda}_0^{(h)} - \tilde{\boldsymbol{U}}^{(h)^T} D(\boldsymbol{\Lambda}_0^{(h)}) e^{\boldsymbol{Z}\boldsymbol{\beta}^{(h)}} + \pi_\lambda(\boldsymbol{\eta}_\lambda|\sigma_{\eta_\lambda}^{2\;(h)}) \; .$$

Candidate generation may depend on the parametric form, but we have found that if the distributions are parametrized in such a way that their parameters are unconstrained, multivariate Gaussian transition kernels yield good results. Effective parametrizations are discussed in appendix 2.B.

2.2.3.6 Updating frailty density parametric component parameters η_u

The loglikelihood for the parameters corresponding to the parametric component of the baseline hazard is given by

$$\ell\left(\boldsymbol{\eta}_u|\ldots\right) = \mathbf{1}_m^T \log \boldsymbol{f}^{(h)} + \pi_u(\boldsymbol{\eta}_u|{\sigma^2_{\eta_u}}^{(h)})$$

Again, candidate generation is discussed in appendix 2.B.

2.2.3.7 Updating the weights ϕ_λ and ϕ_u

The relative weights of the parametric and nonparametric components for the baseline hazard and frailty curves have the following likelihoods:

$$\begin{aligned}
\ell\left(\phi_\lambda|\ldots\right) &= \boldsymbol{\delta}^T \log \boldsymbol{\lambda}_0^{(h)} - \tilde{\boldsymbol{U}}^{(h)^T} D(\boldsymbol{\Lambda}_0^{(h)}) e^{\boldsymbol{Z}\boldsymbol{\beta}^{(h)}} \\
&\quad +(\alpha_{\phi_\lambda 1}-1)\log\phi_\lambda^{(h)} + (\alpha_{\phi_\lambda 2}-1)\log(1-\phi_\lambda^{(h)}) \\
\ell\left(\phi_u|\ldots\right) &= \mathbf{1}_m^T \log \boldsymbol{f}^{(h)} + (\alpha_{\phi_u 1}-1)\log\phi_u^{(h)} + (\alpha_{\phi_u 2}-1)\log(1-\phi_u^{(h)})
\end{aligned}$$

We generate candidates for $\tilde{\phi}_\lambda^{(h+1)}$ using a Beta transition kernel with mean $\phi_\lambda^{(h)}$ and variance $\gamma_{\phi\lambda}$, where the latter is a tuning parameter, and analogously for $\tilde{\phi}_u^{(h+1)}$.

2.2.3.8 Generate the error variance parameters σ^2

Parameters $(\sigma^2_\beta, \sigma^2_\lambda, \sigma^2_u)$ are sampled from the following inverse-gamma distributions:

$$\begin{aligned}
\sigma^2_\beta &\sim IG\left(\frac{p}{2} + \alpha_{\beta 1}, \frac{\boldsymbol{\beta}^{(h)^T}\boldsymbol{\beta}^{(h)}}{2} + \alpha_{\beta 2}\right) \\
\sigma^2_\lambda &\sim IG\left(\frac{K^{(h)}_\lambda}{2} + \alpha_{\lambda 1}, \frac{p_\lambda(\boldsymbol{\theta}^{(h)}_\lambda)}{2} + \alpha_{\lambda 2}\right) \\
\sigma^2_u &\sim IG\left(\frac{K^{(h)}_u}{2} + \alpha_{u 1}, \frac{p_u(\boldsymbol{\theta}^{(h)}_u)}{2} + \alpha_{u 2}\right)
\end{aligned}$$

For the remaining parameters corresponding to the parametric components, other priors may be appropriate depending on the parametric form and parametrization chosen. We discuss these in appendix 2.B.

2.2.3.9 Setting Metropolis-Hastings tuning parameters

The preceding steps in Section 2.2.3.1–2.2.3.8 depend on tuning parameters $\gamma_\nu, \gamma_\beta, \gamma_\lambda, \gamma_u, \gamma_{\eta\lambda}, \gamma_{\eta_u}, \gamma_{\phi\lambda}, \gamma_{\phi_u}$, which must be set in such a way that the acceptance rate of each of the Metropolis-Hastings steps is approximately 25%. It is infeasible to calibrate so many parameters by hand, so we offer the following heuristic:

During the burn-in phase of length B iterations, the MCMC loop of Section 2.2.3.1–2.2.3.8 and possibly Section 2.2.4 may be interrupted every b iterations, $b < B$. The tuning parameters and acceptance rates used during each previous interval of length b can then be used to predict the values of the tuning parameters for which the acceptance rates are 25%, e.g. using linear regression, and the results can be used as tuning parameter values for the next b iterations.

The value of b should be chosen so that B/b is sufficient to yield a large number of evaluations. After the end of the burn-in, the tuning parameters are held fixed.

Although simplistic, we have found that this method works very well, and yields acceptance rates that are very close to 25%.

2.2.4 Reversible-Jump MCMC for adaptive knot selection

In all the steps discussed in Section 2.2.3, the number of knots in the model, and hence the dimension of the spline parameters $\boldsymbol{\theta}_\lambda, \boldsymbol{\theta}_u$, has remained fixed. In order to enable adaptive knot selection, we not only allow knots to move, but also permit changes in dimension, such as adding a knot (birth step) or deleting a knot (death step).

We discuss the procedure in general terms only, since it is identical for the hazard spline and the frailty density spline, and we omit subscripts that identify the parameters as referring to either curve. As before, let $N^{(h)}$ denote the number of interior spline knots $\boldsymbol{\xi}^{(h)}$, and $\boldsymbol{\theta}^{(h)}$ the spline parameter vector of length $K^{(h)} = N^{(h)} + Q$, at iteration h. Let $\pi_N(n)$ denote the prior on the number of knots.

As detailed in Green (1995), changes in model dimension in reversible-jump MCMC are subject to a "dimension–matching" constraint. Typically, transitions between a model indexed by a parameter set $\boldsymbol{\theta}$ of dimension k and a candidate model indexed by parameters $\tilde{\boldsymbol{\theta}}$ of dimension $\tilde{k}$ are accomplished by generating m uniform random numbers $\boldsymbol{u}$ and computing the candidate by a deterministic function $\tilde{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}(\boldsymbol{\theta}, \boldsymbol{u})$. For the reverse move, one generates $\tilde{m}$ random numbers $\tilde{\boldsymbol{u}}$ and computes the candidate as $\boldsymbol{\theta} = \boldsymbol{\theta}(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{u}})$. To ensure reversibility, the mapping

between $(\boldsymbol{\theta}, \boldsymbol{u})$ and $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{u}})$ must be bijective, and in particular, the dimension–matching constraint $m + k = \tilde{m} + \tilde{k}$ must hold.

In our context, adaptive knot selection requires three types of steps: the "move" step, in which the position of a single knot is changed to some new point between its neighbor knots, the "birth" step, in which a new knot is added after a randomly chosen knot and the dimension of the spline parameter $\boldsymbol{\theta}$ increases, and the "death" step, in which a randomly chosen knot is removed and the dimension of the parameter decreases. The move step requires no dimension change, and Metropolis-Hastings methods are sufficient. The death and birth steps however are subject to the reversibility and dimension-matching constraints.

Following Denison et al. (1998), at each iteration we choose randomly whether to execute a birth, death, or move step. Given that $N_\lambda^{(h)} = n$, and the probabilites b_n, d_n, m_n of birth, death and move steps respectively are set to:

$$b_n = c \min\left\{1, \frac{\pi_N(n+1)}{\pi_N(n)}\right\}, \quad d_n = c \min\left\{1, \frac{\pi_N(n-1)}{\pi_N(n)}\right\}, \quad m_n = 1 - b_n - d_n ,$$

where the constant c controls the rate of dimension-changing steps, and is set to $c = 0.4$ as in Denison et al. (1998). These parameters are chosen so that

$$b_n \pi_N(n) = d_{n+1} \pi_N(n+1) . \tag{2.18}$$

We give details on the move step in Section 2.2.4.1, the birth step in Section 2.2.4.2, and the death step in Section 2.2.4.3.

2.2.4.1 Knot position change (move step)

In the move step, a single knot position $\xi_k^{(h)}$ to be moved is chosen uniformly from the set of interior knots, and changed to a random new candidate position located

between its neighboring knots. That is, the candidate knot position $\tilde{\xi}_k^{(h)}$ is selected uniformly from the set of candidate locations $\xi^c \in \boldsymbol{\xi}^c$ such that $\xi_{k-1}^{(h)} < \xi^c < \xi_{k+1}^{(h)}$. The spline parameters $\boldsymbol{\theta}$ remain unchanged.

Since the prior on the knot positions is uniform over the set of candidate knots, the priors for knots $\boldsymbol{\xi}^{(h)}$ and the candidate $\tilde{\boldsymbol{\xi}}^{(h+1)}$ are identical. Since no dimension change is required, the new knot positions are accepted with probability

$$\rho = \min\left\{1, \frac{\mathcal{L}(\boldsymbol{\theta}|\tilde{\boldsymbol{\xi}}^{(h+1)}, \ldots)}{\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\xi}^{(h)}, \ldots)}\right\},$$

where $\mathcal{L}$ is posterior spline parameter likelihood given in eq. (2.16) or eq. (2.17) for the hazard and frailty spline respectively. Note that these likelihoods depend on the knot positions through $\boldsymbol{\lambda}_0^{(h)}$ and $\boldsymbol{f}^{(h)}$.

2.2.4.2 Knot addition (birth step)

In the birth move, a random unoccupied candidate knot $\xi^c \in \boldsymbol{\xi}^c$ is chosen to be added to the current set of knots $\boldsymbol{\xi}^{(h)}$, of length $N^{(h)}$. Denote by k the interval of the current knot set containing ξ^c, so that $\xi_k^{(h)} < \xi^c < \xi_{k+1}^{(h)}$. The new candidate knot set is then given by

$$\tilde{\boldsymbol{\xi}}^{(h+1)} = \left\{\xi_1^{(h)}, \ldots, \xi_k^{(h)}, \xi^c, \xi_{k+1}^{(h)}, \ldots, \xi_{N^{(h)}}^{(h)}\right\},$$

of length $\tilde{N}^{(h+1)} = N^{(h)} + 1$.

The set of spline coefficients $\boldsymbol{\theta}^{(h)}$ of length $K^{(h)}$ must be updated to a candidate set $\tilde{\boldsymbol{\theta}}^{(h+1)}$ of length $\tilde{K}^{(h+1)} = K^{(h)} + 1$. There are simple rules for non-destructively inserting a new knot into a B-spline function (de Boor, 2001), but using these directly would violate the reversibility and dimension-matching constraint between the birth and death moves mentioned earlier. Since the birth move begins in a

model of dimension $K^{(h)}$ and its reverse begins at dimension $K^{(h)}+1$, we need to generate an additional random number for the birth move. Intuitively, since removing a knot is a destructive procedure and may cause the shape of the curve to change, we must during the birth move be able to generate the set of curves that would reduce to the original curve upon removal of the new knot.

To do this, we compute the candidate spline parameters $\tilde{\boldsymbol{\theta}}^{(h+1)}$ for inserting a knot $\xi^c \in (\xi_k^{(h)}, \xi_{k+1}^{(h)})$ as follows:

$$\tilde{\theta}_j^{(h+1)} = \begin{cases} \theta_j^{(h)} & \text{if } j \leq k+1 \\ \theta_{j-1}^{(h)} & \text{if } j > k+Q \\ \log\left(r_j e^{\theta_j^{(h)}} + (1-r_j)e^{\theta_{j-1}^{(h)}}\right) & \text{if } k+1 < j < k+Q \\ \log\left(u e^{\theta_j^{(h)}} + (1-u)e^{\theta_{j-1}^{(h)}}\right) & \text{if } j = k+Q \end{cases} \tag{2.19}$$

where $r_j = (\xi^c - \xi_{j-Q}^{(h)})/(\xi_{j-1}^{(h)} - \xi_{j-Q}^{(h)})$, and $u \sim U(0,1)$. These rules correspond to the deterministic rules in de Boor (2001), except that the parameter $\tilde{\theta}_{k+Q}^{(h+1)}$ is perturbed by a random amount, rather than by the knot ratio r_{k+Q}.

The prior ratio for the birth move is given by product of the ratio of the priors on the number of knots, the priors on the knot positions, and the priors on the spline parameters:

$$R_P = \frac{\pi_N(N^{(h)}+1)}{\pi_N(N^{(h)})} \cdot \frac{N^{(h)}+1}{M-N^{(h)}} \cdot \frac{\pi(\tilde{\boldsymbol{\theta}}^{(h+1)}|\sigma^2)}{\pi(\boldsymbol{\theta}^{(h)}|\sigma^2)} \ ,$$

since the prior on the knot positions is that they are randomly chosen among the candidate knots, so that $\pi(\boldsymbol{\xi}^{(h)}) = \left[\binom{M}{N^{(h)}}\right]^{-1}$. The transition ratio is given by

$$R_T = \frac{d_{N^{(h)}+1}/(N^{(h)}+1)}{b_{N^{(h)}}/(M-N^{(h)})} \ .$$

Note that together with eq. (2.18), this implies that

$$R_P \cdot R_T = \frac{\pi(\tilde{\boldsymbol{\theta}}^{(h+1)}|\sigma^2)}{\pi(\boldsymbol{\theta}^{(h)}|\sigma^2)} = (2\pi\sigma)^{-\frac{1}{2}} \exp\left(\frac{p(\tilde{\boldsymbol{\theta}}^{(h+1)}) - p(\boldsymbol{\theta}^{(h)})}{2\sigma^2}\right) \ .$$

The likelihood ratio R_L is given by the ratios of the likelihoods for the spline parameters $\boldsymbol{\theta}^{(h)}$, whose logarithms are given by either eq. (2.16) or eq. (2.17), without the prior penalty terms. Lastly, the Jacobian for the transformation from $(\boldsymbol{\theta}^{(h)}, u)$ to $(\tilde{\boldsymbol{\theta}}^{(h+1)})$ in eq. (2.19) is

$$|J| = \left| \frac{\left(\exp(\theta_{k+Q}^{(h)}) - \exp(\theta_{k+Q-1}^{(h)})\right)}{\exp(\tilde{\theta}_{k+Q}^{(h+1)})} \prod_{j=k+2}^{k+Q-1} \frac{r_j \cdot \exp(\theta_j^{(h)})}{\exp(\tilde{\theta}_j^{(h+1)})} \right|$$

The candidate number of knots $\tilde{N}^{(h+1)}$ and spline parameters $\tilde{\boldsymbol{\theta}}^{(h+1)}$ are then accepted with probability

$$\rho = \min\left\{1, R_L \cdot R_P \cdot R_T \cdot |J|\right\} \ . \tag{2.20}$$

2.2.4.3 Knot deletion (death step)

In the death step, a single knot $\xi_k^{(h)}$ is chosen uniformly from the set of knots $\boldsymbol{\xi}^{(h)}$ to be removed. The candidate knot set for the next iteration is then $\tilde{\boldsymbol{\xi}}^{(h+1)} = \left\{\xi_1^{(h)}, \ldots, \xi_{k-1}^{(h)}, \xi_{k+1}^{(h)}, \ldots, \xi_{N^{(h)}}^{(h)}\right\}$. The spline parameters are correspondingly adjusted by the inverse of the transformation in eq. (2.19), that is, by deleting the parameter $\theta_{k+Q-1}^{(h)}$ and adjusting the remaining parameters as

$$\tilde{\theta}_j^{(h+1)} = \begin{cases} \theta_j^{(h)} & \text{if } j < k+1 \\ \theta_{j+1}^{(h)} & \text{if } j \geq k+Q-1 \\ \log\left(\frac{1}{r_j} e^{\theta_j^{(h)}} - \frac{1-r_j}{r_j} e^{\theta_{j-1}^{(h)}}\right) & \text{if } k+1 \leq j < k+Q-1 \end{cases} \tag{2.21}$$

Because the birth and death moves are symmetrically defined, the likelihood ratio, prior ratio, transition ratio and Jacobian determinant are the inverses of those in eq. (2.20).

2.3 Simulation Studies

We implemented the methodology described in Section 2.2 in the `R` package `splinesurv`. In order to establish the performance and flexibility of the method, we conducted simulation studies under a variety of settings.

Our simulations investigate the capacity of the method to correctly identify the form of the underlying baseline hazard and frailty density, as a function of the number of clusters and cluster size. Furthermore, we wish to show that the method can be used to accurately estimate the regression coefficients $\boldsymbol{\beta}$ and the frailty variance, which we will henceforth denote by σ^2.

We consider three scenarios within which to test the method, differing in the form of the "true" baseline hazard and frailty density used to generate simulated data. The first, referred to as the "Parametric" scenario, is characterized by a Weibull hazard of scale 1 and shape 1.8, and lognormal frailty density with variance .25, both of which are standard forms typically well-modeled by parametric methods. In the second scenario, referred to as the "Smooth" scenario, the baseline hazard is a smooth curve that cannot be well-described by typical parametric forms, and the frailty density is a mixture of two lognormal distributions. In third and final scenario, referred to as the "Stepfunction" scenario, the baseline hazard is a discontinuous step function, and the frailty distribution is a mixture of uniforms. Figure 2.2 contains plots of the hazard and frailty in each of the three scenarios.

For purposes of the simulation, a replication consists of first generating frailties U_i, $i = 1 \ldots m$ from the scenario's frailty density. A single covariate is generated for each subject as $Z \sim N(0, 1)$. The single regression coefficient is fixed at $\beta = 1$. Given the frailty and covariate, event times can then be generated using the

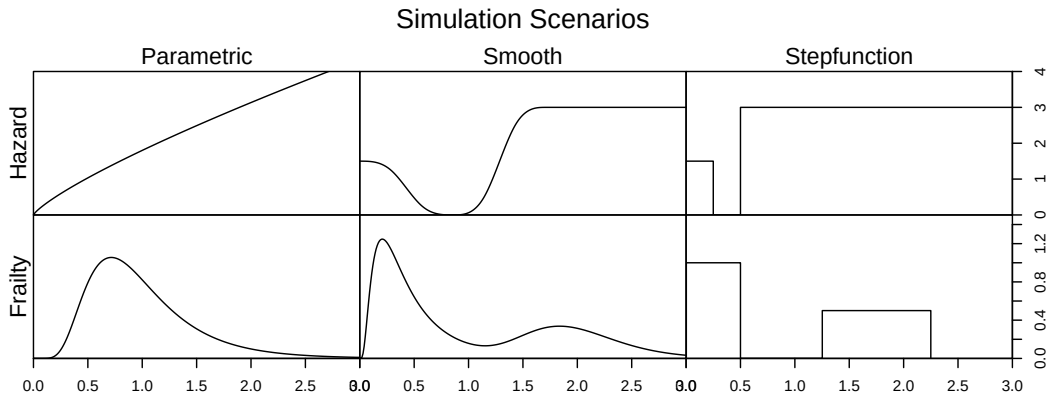

Figure 2.2: "True" baseline hazard curves and frailty densities used for generating simulated data in each of the three simulation scenarios

baseline hazard for the scenario. Censoring times are independently generated from a Weibull hazard with shape $\gamma_C = 1.8$ and scale λ_C chosen for each scenario to yield approximately a 20% censoring rate ($\lambda_C = .15$ in the Parametric scenario, and $\lambda_C = .1$ in the Smooth and Stepfunction scenarios). The sample generated in this way can then be fit using the `splinesurv` package.

2.3.1 Curve fitting performance

To explore the effects of sample size on the quality of the curve fits, we first conduct a single replication for various sample sizes, under each scenario, and explore the effect of different model specifications. We limit ourselves to four sample sizes for each scenario, setting the number of clusters to either $m = 10$ or $m = 500$, and the cluster size to either $J_i = 10$ or $J_i = 500$, $i = 1 \ldots m$.

The methodology is very flexible, and offers a range of choices of penalty functions, parametric distributions, prior parameters, and the option of adaptive knot selection, but for brevity, we only select one model specification for purposes of

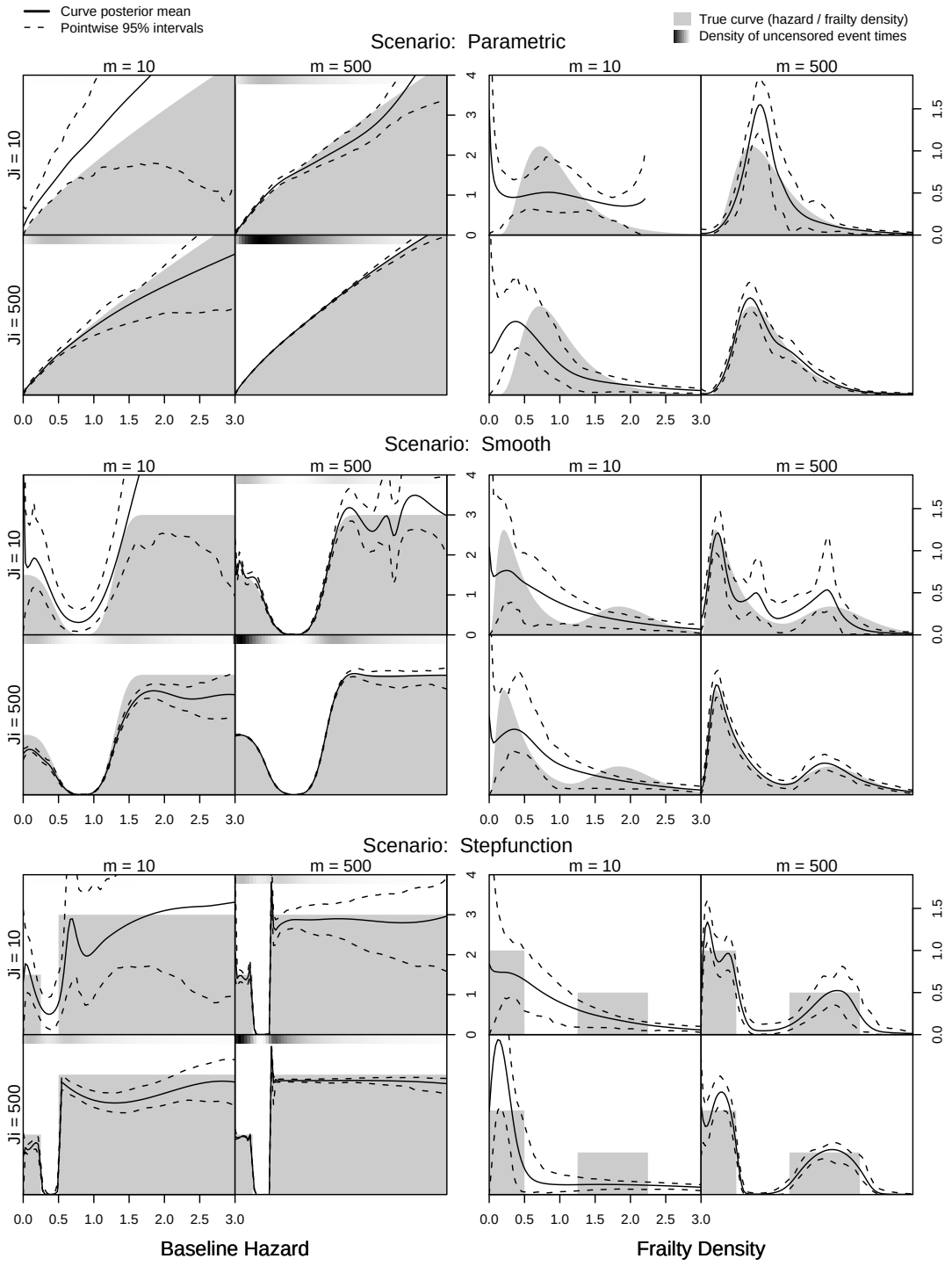

Figure 2.3: Baseline hazard and frailty density curve fitting results of a single replication conducted under each of the three scenarios, for different sample sizes.

demonstrating curve-fitting here. For both the hazard and frailty, we include a spline component only, using a simple Gaussian prior on the spline parameters (corresponding to the penalty function in Section 2.A.1). We allow for adaptive knot selection with a truncated Poisson prior on the number of knots, with means $\mu_\lambda = \mu_u = 10$ and a maximum of 35 knots, and 100 candidate knots distributed uniformly over the range of the data. Each fit was run for a 2000-iteration burn-in, during which tuning parameters were chosen adaptively to ensure approximately a 25% parameter acceptance rate, followed by 3000 iterations used to construct posterior estimates.

The results are shown in Figure 2.3, and indicate that the methodology functions as intended: In all three scenarios, the fitted models capture the features of the underlying hazard and frailty curves with sufficiently large samples. The number of clusters appears to have a more immediate effect on the quality of the fit than the cluster size, especially for the frailty density. In order to obtain an accurate estimate of the frailty density, a large number of clusters is required, but these clusters need not be large. With few, large clusters, the form of the hazard can be identified, although that of the frailty cannot. Hazard estimates in the Stepfunction scenario display sharp spikes at the points of discontinuity—this is an artifact caused by the use of cubic splines in a scenario where linear splines would have been better able to capture the discontinuity.

Further such simulation results (not shown here) indicate that fixed-knot penalized splines perform well in the Parametric and Smooth scenarios, but do quite poorly in the Stepfunction scenario, as the sharp trough cannot be captured without significant smoothing error. Further, the inclusion of a correctly specified

parametric component improves curve-fitting performance in the Parametric scenario, and does not significantly affect the other scenarios.

2.3.2 Parameter estimation performance

In order to establish the ability of the procedure to estimate the regression parameter and frailty variance, we conduct a simulation study at smaller sample sizes. Sample sizes under all three scenarios are limited to 10, 50 and 500 clusters of size 10 or 50, excluding the largest combination. The model and scenarios are specified as before.

We consider four model specifications: The first is a fourth order spline-only model with adaptively chosen knots and a Poisson(10) prior on the number of knots, as described for Figure 2.3. The second includes additionally parametric components, consisting of a Weibull baseline and a lognormal frailty curve, with a Beta$(1, 2)$ prior on the weight, thus giving slight preference to the parametric component. The third is a fixed-knot penalized spline model specified according to Section 2.1.1, with equally spaced knots, and penalties on the squared second differences between the parameters, following Section 2.A.2. The fourth is similar, but penalizes the integrated squared second derivative, as per Section 2.A.3. Penalized spline fits are fairly sensitive to the choice of hyperparameters, so they were chosen here so as to give reasonably smooth curves in several test scenarios; we intentionally did not choose the "best" settings, but instead selected parameters as one might do if the curve were unknown.

In Bayesian estimation by MCMC, the collection of posterior samples contains far more information than the point estimates and intervals generally available in

Table 2.1: Biases and 95% interval coverage probabilities for point estimators of the regression coefficient and frailty variance, compared between a spline-only model using adaptive knot selection, a model which additionally includes a parametric component, two fixed-knot penalized spline models with penalties on the parameter second differences and integrated second derivative respectively, and gamma frailty models fit by the `coxph` package. Results are based on 1000 replications.

				Spline		Spline+Par		Pen. (2^{nd} diff.)		Pen. (2^{nd} deriv.)		Cox: gamma	
Scenario		m	J_i	β	σ^2	β	σ^2	β	σ^2	β	σ^2	β	σ^2
Parametric	Bias (%)	10	10	1.49	50.63	−0.50	20.39	1.98	11.60	3.74	−10.57	0.56	−32.96
			50	0.55	6.56	0.16	−4.62	0.11	−1.85	0.62	−12.73	0.85	13.85
		50	10	2.39	30.22	1.06	12.18	−1.00	−30.01	0.48	−10.88	−0.26	−24.00
			50	0.83	10.39	0.23	1.50	−0.50	−12.13	0.88	−1.24	0.72	35.95
		500	10	0.79	8.82	0.54	5.61	−0.31	3.93	0.61	−1.25	−0.26	−20.65
	95% CP	10	10	95.60	92.80	92.70	91.20	90.49	95.10	85.94	88.65	94.55	
			50	95.60	68.43	94.51	61.94	91.68	65.40	87.84	62.01	94.86	
		50	10	95.20	83.90	92.30	90.30	93.19	68.34	88.22	81.07	93.60	
			50	95.80	69.23	94.21	68.13	93.66	59.15	87.55	61.55	93.96	
		500	10	95.20	89.80	93.70	90.10	91.39	91.59	84.49	86.20	93.80	
Smooth	Bias (%)	10	10	2.40	31.11	−1.59	22.68	−0.06	0.03	−14.33	−36.83	−0.21	1.36
			50	0.98	12.37	−0.20	8.96	0.14	3.23	−4.97	−13.33	0.32	27.26
		50	10	1.59	19.21	0.66	20.03	−0.12	−8.53	−5.05	−15.05	0.19	11.34
			50	0.41	6.40	0.17	5.99	−0.06	−1.36	−1.26	−6.42	0.29	34.54
		500	10	0.63	8.32	0.68	13.05	0.03	3.76	−0.28	−3.62	−0.02	14.52
	95% CP	10	10	93.90	83.10	93.70	83.70	94.29	81.88	79.67	54.44	94.10	
			50	94.81	50.95	94.81	54.15	93.06	54.33	82.00	47.03	94.56	
		50	10	94.90	77.70	94.40	78.80	94.16	74.04	83.83	71.41	94.30	
			50	95.70	54.10	95.90	54.60	94.14	45.70	88.79	51.25	95.05	
		500	10	94.10	72.70	94.30	50.90	94.60	86.30	88.11	93.12	93.45	
Stepfunction	Bias (%)	10	10	−0.37	45.23	−3.36	35.72	−0.81	20.30	−9.49	−24.22	0.70	61.01
			50	0.95	21.66	−0.34	25.02	−0.32	16.98	−5.10	−0.87	0.53	74.14
		50	10	0.40	31.18	−0.11	36.85	−0.28	5.34	−5.25	−7.58	0.17	68.63
			50	0.03	9.94	0.11	9.32	−0.31	3.50	−1.89	0.24	0.04	72.00
		500	10	0.55	12.45	0.68	16.43	−0.28	8.81	−0.58	6.15	0.28	70.51
	95% CP	10	10	95.70	72.00	93.60	74.70	92.69	76.55	88.41	66.26	94.55	
			50	94.71	30.37	94.71	30.07	93.47	31.43	85.54	38.09	94.46	
		50	10	94.89	51.95	95.20	42.30	94.27	65.03	84.11	72.40	95.05	
			50	95.60	34.07	96.60	30.97	94.16	29.20	85.71	29.60	94.71	
		500	10	94.50	46.50	95.50	30.23	94.77	65.93	90.29	68.10	93.55	

frequentist methods. For purposes of the simulation study, and to enable comparison with frequentist methodology, we construct point estimators based on the posterior samples. A natural estimator for the regression coefficient β is the posterior mean of its distribution, estimated by the sample average of MCMC samples. For the frailty variance σ^2, one natural estimator is the sample variance of the density functions constructed from the MCMC samples of the spline parameters $\boldsymbol{\theta}_u$. A second estimator is variance of the frailty samples U_i, averaged over all iterations. In practice, we found that the latter performs slightly better than the former, since it is less directly affected by smoothing bias.

Table 2.1 contains estimates of the biases of these point estimators, based on 1000 simulations, with the four model specifications, and an extended Cox model with gamma frailties as described in Therneau and Grambsch (2000), fitted using the routine `coxph` for comparison.

Results show that the model yields good estimates of the regression coefficient, particularly for larger samples. In the Parametric and Smooth scenarios, the Penalized (2^{nd} diff.) model has the lowest regression parameter bias for large samples. Due to the sensitivity of penalized model fits to hyperparameters, the quality of these results cannot be accurately judged, as the choice of hyperparameters may have been particularly fortuitous. The Penalized (2^{nd} deriv.) model often underestimates the regression coefficient and frailty variance. These results suggest that if enough prior information about the process exists to make reasonable choices about the hyperparameters, a penalized spline model can be a good choice.

An advantage of the adaptive knot selection method is significantly lower sensitivity to hyperparameters. Since the adaptive method allows the smoothness of the spline to be controlled through the prior on the number of knots rather than

through an explicit penalty, there are fewer settings that need to be manually adjusted. The spline-only method performs well, particularly with larger clusters. When clusters are small, the method tends to overestimate the frailty variance, because smoothing error and uncertainty affect the variance estimates. This effect is particularly severe in the Smooth and Stepfunction scenarios—this agrees with the oversmoothing observable in Figure 2.3 for smaller samples. Including a parametric component has a beneficial effect on the estimates in the Parametric scenario, improving both regression coefficient and frailty estimates, but does not have significant detriment in other scenarios.

2.4 Data Examples

We illustrate the use of the proposed methodology with two example data sets. The first is a set of observations of congestive heart failure patients gathered in the course of a randomized clinical trial, which we reanalyze in Section 2.4.1 with the secondary goal of identifying the effect of various factors on the risk of rehospitalization or death. The second is a study of diabetic retinopathy analyzed multiple times in the statistical literature, including by Huster et al. (1989) and Therneau and Grambsch (2000), used to illustrate the effects of adaptive knot selection and penalized smoothing in Section 2.4.2.

2.4.1 Congestive heart failure data

The study was conducted in a 487-bed, not-for-profit community hospital located in southeast Michigan. The study population consisted of patients with either systolic or diastolic heart failure assembled for the original purpose of a randomized,

controlled trial comparing a pro-active case management strategy versus usual care on all-cause re-hospitalizations. A planned secondary analysis was to determine prognostic factors for readmission or mortality.

Patients were eligible for the study if they were hospitalized on an internal or family medicine service between October 29, 2002 and September 20, 2003 and received intravenous diuretics to treat possible heart failure. Intervention patients were assessed by a cardiology nurse practitioner who developed a protocol-driven discharge plan that could include telemanagement, an outpatient nurse-run heart failure clinic, or usual care. All control patients were managed by the usual discharge planning activities of hospital staff.

Computations based on pilot data adjusted for the impact of clustering suggested a sample size of 440 patients for the study. Of these, 17 died during the index hospitalization and were removed from the sample, resulting in a cohort of 423 patients. Unfortunately, half the patients assigned to the intervention arm were discharged prior to receiving the complete intervention, and the study could not be completed as planned.

Using the intention-to-treat approach, no difference between the intervention or control groups was found for the outcome of all-cause subsequent hospitalizations or emergency department encounters.

We here proceed to re-analyze the data, defining the event of interest as a patient's rehospitalization or death during the 180 day period following the index hospitalization. 257 such events were observed, of which 233 are rehospitalizations and the remainder are deaths. The remaining 39.2% of patients are treated as censored observations at the end of followup.

Table 2.2: Covariates and basic descriptive statistics for the congestive heart failure data

Name	Description	Mean	Median	SD
hxsumINPTorER	Prior hospitalizations and ED visits (count)	0.88	0.00	1.31
minHb	Minimum hemoglobin	10.79	10.80	2.07
LN_lastCREAT	Last creatinine log	0.26	0.22	0.45
LN_maxGLU	Maximum glucose log	5.12	5.04	0.41
minPLTSlt50k	Indicator: minimum platelet count ¡10^5	0.02	0.00	0.13
lastPOTASgt5	Indicator: last potassium >5	0.04	0.00	0.19
itoECF	Indicator: discharged to nursing home	0.19	0.00	0.39
ejectionpctcon	Cardiac ejection fraction	43.66	45.00	16.47
dcbeta	Indicator: beta-blockers	0.54	1.00	0.50
dcaceiorarb	Indicator: ACE Inhibitors	0.61	1.00	0.49

Patients are clustered into 31 groups by their attending physician, ranging in size between 1 and 80 patients, with mean and median cluster size of 14 and 5.5 respectively.

A wide range of explanatory variable data are available for each patient. For purposes of the analysis, covariates with more than 5% missing values were removed from the data set, and the remaining missing values were imputed with the median. Prior to analysis, all covariates were centered and standardized. We found experimentally that doing so improved the mixing properties of the MCMC procedure. A subset of covariates was selected by a combination of stepwise automated procedures and consultations with the study clinicians, and is shown together with basic descriptive statistics in Table 2.2. Since the treatment could not be administered to to half of the intervention group patients, treatment group membership was excluded from the set of covariates.

We first fit a Spline-Only model by specifying the hazard and frailty following Sections 2.1.1 and 2.1.3, as fourth-order (cubic) splines with truncated Poisson priors on the number of knots with mean 20 and a maximum of 35 knots, and

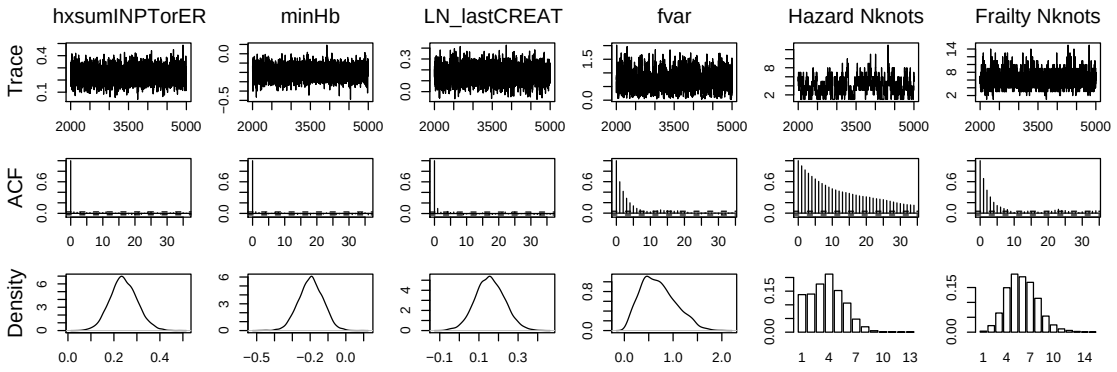

Figure 2.4: Trace plots, autocorrelation functions, and posterior density estimates for a Spline-Only fit to the regression coefficients of the congestive heart failure data. Plots are shown for three of the regression coefficients, the frailty variance, and the number of spline knots for the hazard and frailty.

run the chain for 50,000 iterations, discarding the first 20,000 as burn-in and thinning the chain to every 10th sample. Table 2.3 shows the posterior mean and 95% posterior intervals for the covariate effects and the variance of the random effect in the first three columns. There is substantial agreement on the signs and magnitudes of the coefficients with gamma and lognormal frailty models fitted by `coxph`, although notably, the gamma frailty model estimates the frailty variance as zero, and the lognormal model's frailty variance estimate is very small.

We monitor the mixing of chain parameters by examining trace plots and autocorrelation functions of the posterior samples. The trace plots for coefficients in Figure 2.4 indicate that the regression coefficient estimates have converged, and that the degree of thinning is adequate, and kernel density estimates based on the posterior samples suggest approximately normal posterior distributions. Estimates of the frailty variance and number of spline knots mix at a considerably lower rate.

The top panel of Figure 2.5 shows the posterior mean estimate of the hazard, survival and frailty density curves, as well as pointwise 95% credible bands for

Table 2.3: Posterior means and 95% credible intervals of regression coefficients and frailty variance, for a Spline-Only model, and a model additionally including a Weibull/Lognormal parametric component, fitted to the congestive heart failure data, with gamma and lognormal frailty model estimates from `coxph` for comparison.

	Spline Only			Spline + Parametric		
Covariate	PM	2.5%	97.5%	PM	2.5%	97.5%
hxsumINPTorER	0.242	0.128	0.365	0.251	0.137	0.364
minHb	−0.196	−0.329	−0.067	−0.197	−0.334	−0.066
LN_lastCREAT	0.156	0.014	0.305	0.168	0.029	0.309
LN_maxGLU	0.112	−0.009	0.235	0.115	−0.013	0.241
minPLTSlt50k	0.120	0.007	0.223	0.120	0.006	0.223
lastPOTASgt5	0.078	−0.045	0.186	0.083	−0.040	0.191
itoECF	0.101	−0.020	0.222	0.106	−0.010	0.222
ejectionpctcon	−0.102	−0.231	0.029	−0.111	−0.245	0.019
dcbeta	−0.039	−0.159	0.086	−0.040	−0.163	0.085
dcaceiorarb	−0.037	−0.167	0.088	−0.034	−0.162	0.090
minHb:LN_lastCREAT	0.165	0.031	0.305	0.165	0.037	0.312
Frailty Variance	0.697	0.131	1.443	0.439	0.038	1.395

	Cox: gamma			Cox: lognormal		
Covariate	Est	SD	Pval	Est	SD	Pval
hxsumINPTorER	0.248	0.056	0.000	0.249	0.056	0.000
minHb	−0.193	0.069	0.005	−0.195	0.070	0.005
LN_lastCREAT	0.236	0.074	0.001	0.235	0.074	0.002
LN_maxGLU	0.122	0.067	0.068	0.123	0.067	0.066
minPLTSlt50k	0.138	0.051	0.007	0.138	0.051	0.007
lastPOTASgt5	0.122	0.055	0.026	0.120	0.055	0.028
itoECF	0.129	0.060	0.033	0.129	0.060	0.033
ejectionpctcon	−0.117	0.071	0.098	−0.119	0.071	0.094
dcbeta	−0.055	0.065	0.398	−0.055	0.065	0.400
dcaceiorarb	−0.045	0.066	0.502	−0.043	0.066	0.517
minHb:LN_lastCREAT	0.189	0.073	0.010	0.190	0.073	0.009
dcaceiorarb	−0.037	−0.167	0.088	−0.034	−0.162	0.090
Frailty Variance	0.000			0.001		

each. The shape of the hazard may inform the timing and duration of future interventions to reduce readmissions and mortality from heart failure. In this dataset, the risk for readmission or death was greatest shortly after discharge from the index hospitalization. The hazard declined rapidly during the first few weeks post-discharge then declined more slowly before reaching a plateau at approximately 120 days. This suggests interventions to postpone mortality or readmissions need

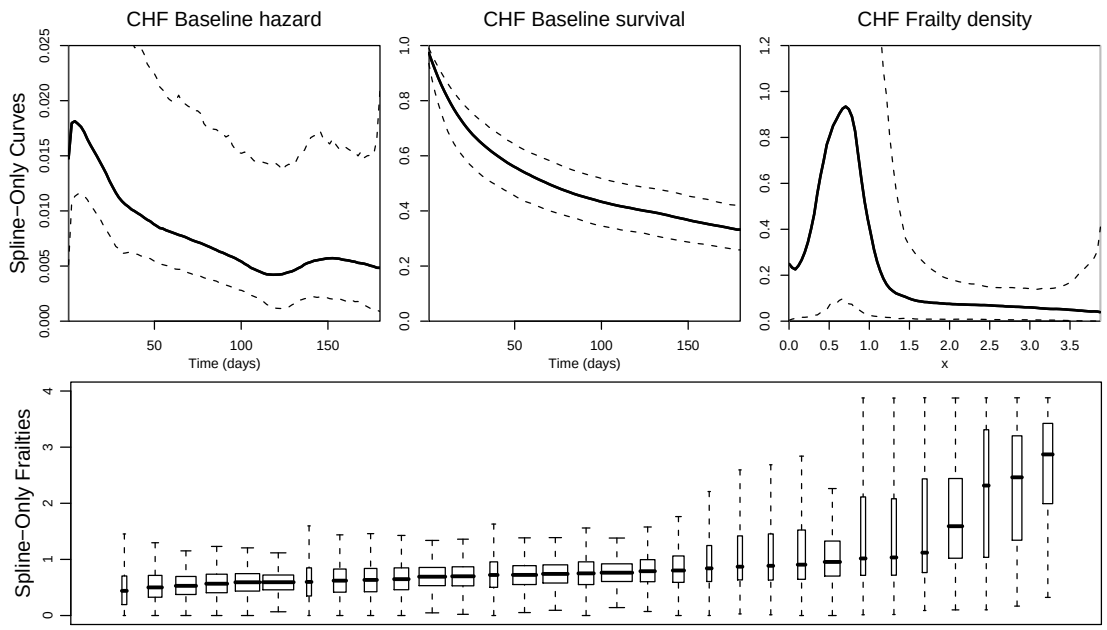

Figure 2.5: Upper panel: Hazard, survival and frailty density estimates, and 95% pointwise posterior intervals for the congestive heart failure data. Spline-Only specification, with adaptive knot selection and a Poisson(20) prior on the number of knots. The baseline represents a patient with average covariates. Lower panel: Boxplots of posterior frailty estimates for each of the 31 clusters, sorted in order of increasing posterior means. Box width indicates the cluster size.

to be targeted at the care transition from the hospital to home or to another facility. Although determining the optimal duration and intensity of the intervention would require formal cost-effectiveness analyses, the hazard curve suggests most of the benefit would be realized within the first few months after the index event.

The Bayesian approach allows the full posterior distributions of the frailties to be examined. The lower panel of Figure 2.5 shows boxplots of the posterior frailty estimates for each of the clusters, to give some indication of the marginal posterior frailty distributions. The difference between the smallest and largest cluster frailties suggests that the large estimated frailty variance is a significant effect, and not merely a product of smoothing error. Frailty estimates for larger

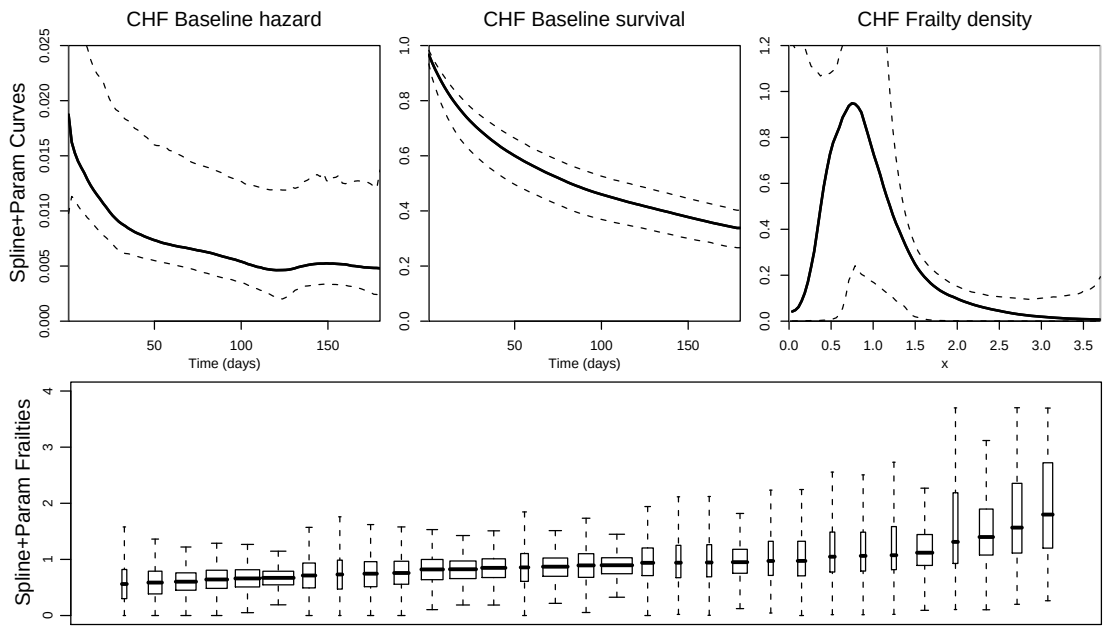

Figure 2.6: Upper panel: Hazard, survival and frailty density estimates, and 95% pointwise posterior intervals for the congestive heart failure data. Spline+Parametric model specification, with the spline component specified as in Figure 2.5 and the Weibull and lognormal parametric components for the hazard and frailty density respectively. Component weight prior is Beta$(1, 2)$, which favors the parametric component. The baseline represents a patient with average covariates. Lower panel: Boxplots of posterior frailty estimates for each of the 31 clusters, sorted in order of increasing posterior means. Box width indicates the cluster size.

clusters are more precise, and small frailty values are estimated more precisely than large ones. The size of the errors relative to the frailties nevertheless suggests that estimation error contributes to the frailty variance in Table 2.3, and it has likely been overestimated.

Since the frailty density appears similar to a Lognormal density, and the hazard may be well-modeled by a Weibull hazard function, we construct a second fit, including parametric components in both curves, following Section 2.1.2. The weights ϕ of the spline component are given Beta$(1, 2)$ priors, which are triangular

priors giving more weight to the parametric component. The posterior means and quantiles in the second set of three columns of Table 2.3 contain similar results for the regression coefficients, but a much smaller posterior mean frailty variance estimate. The curves in Figure. 2.6 show that including a parametric component has a smoothing effect on the hazard and frailty density estimates. In particular, the parametric component dominates for the frailty density, but has a smaller effect on the estimated hazard curve. The lower panel in Figure 2.6 shows that including a parametric component slightly increases the precision with which frailties are estimated, relative to the estimates of Figure 2.5. The effect is particularly noticeable with smaller clusters.

We next use the congestive heart failure data to illustrate the effect of choosing different priors on the number of knots. Figures 2.5 used a Poisson(20) prior on the number of knots. The top panel of Figure 2.7 compares this fit to those resulting from different Poisson prior choices. As expected, setting the prior to Poisson(1) leads to excessively smooth fits, whereas the Poisson(50) curve is considerably more variable and shows potentially undesirable detail. The effect on the survival curve is relatively small, however, as local bumps in the hazard are smoothed out by the integration.

The second and third panels show the effect of using geometric and negative binomial priors for the number of knots. As noted by Biller (2000), these priors universally encourage smoother fits, and are relatively insensitive to the choice of parameters. This is a desirable property if a more robust fit is preferred, but if control over the smoothness of the curve is desired, the Poisson prior is preferable.

The relative prominence of the parametric and spline components can be controlled through the prior on the weights ϕ. The bottom panel of Figure 2.7 shows

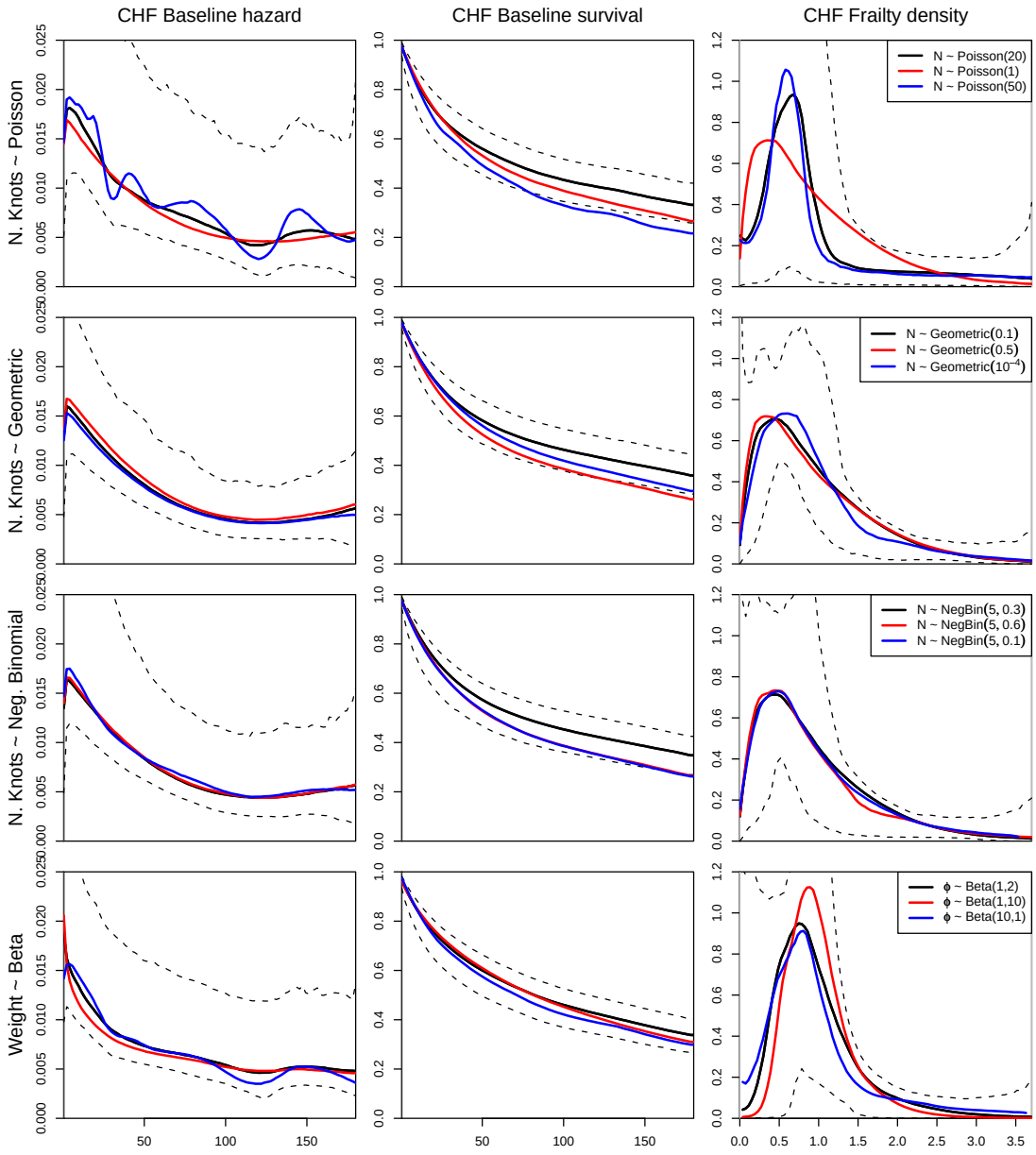

Figure 2.7: Illustrating the effect of choosing different priors on the number of knots, and different prior parameters for the spline weight. Credible intervals shown are for the black line in each plot.

the effects of changing the prior to Beta$(1, 10)$ and Beta$(10, 1)$, which respectively place more and less emphasis on the parametric component. The frailty density is more sensitive to changes in the prior weight than the hazard curve, because with only 31 clusters, the data contains relatively little information about the frailty density.

2.4.2 Diabetic retinopathy data

The data (described in detail in Huster et al. (1989), and in Therneau and Grambsch (2000)) consists of 197 patients with diabetic retinopathy in both eyes, observed during a six-year period to study the effectiveness of a laser photocoagulation treatment in reducing the incidence of blindness. For each patient, one eye was treated, while the other remained untreated as a control, and the time to blindness was measured for each eye. We therefore treat the data as 394 observations in 197 clusters of size 2, allowing the frailty to capture the dependence between the two measurements on each patient.

In addition to the treatment effect, the study also distinguishes between juvenile- and adult-onset diabetes, with 42% of subjects falling into the latter class. We include the treatment effect and onset category as covariates, as well as an interaction term between the two, which was shown to be significant by Huster et al. (1989). We initially fit a cubic spline model with the same settings as the Spline-Only model of Section 2.4.1.

Figure 2.8 shows the estimated hazard and survival probability for each of the four groups in the sample. The adaptive knot selection identifies a spike in the hazard at 13 months, accompanied by a sharp decline in survival, caused by

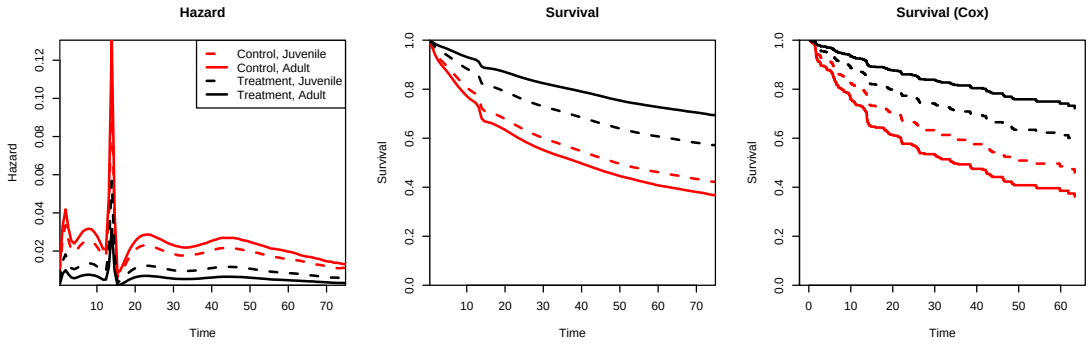

Figure 2.8: Hazard and survival estimates for the four groups in the diabetic retinopathy data for a model with adaptive knot selection and a Poisson(20) prior for the number of knots, and survival curves from a fitted Cox model with gamma frailties for comparison.

a particularly large number of events near that time. This effect is also clearly noticeable in Kaplan-Meier survival function estimates of the data (Huster et al., 1989), and in the fitted Cox model estimates in the rightmost panel of Figure 2.8. In practice, in the absence of a medical explanation for this spike, one might consider the spike to be noise, and wish to smooth the hazard. This can be accomplished within the adaptive knot selection framework by using a different prior on the number of knots, or including a smooth parametric component with a favorable prior, as illustrated for the congestive heart failure data in Figure 2.7 and Figure 2.6. Alternatively, the smoothness of the spline can be controlled explicitly, by disabling adaptive knot selection and relying entirely on penalized smoothing instead: as noted in Section 2.1.1, the prior on the spline parameters in eq. (2.5) and (2.6) may contain a penalty term that produces a smoothing effect.

Table 2.4 shows regression coefficient and frailty variance point estimates for four spline fits: the first is the adaptive fit shown in 2.8, which uses a Poisson(20) prior for the number of knots. For the second, the number of knots is given a Geometric(0.1) prior. The third has 20 fixed knots, spaced equally, and uses a

Table 2.4: Posterior means and 95% credible intervals for fits to the diabetic retinopathy data, (1) using splines with adaptive knot selection, and a Poisson(20) prior for the number of knots, (2) with a Geometric(0.1) prior on the number of knots, (3) with fixed knots and a penalty on the squared second differences on spline parameters, (4) with fixed knots and a penalty on the integrated squared second derivative, (5) a Cox model with gamma frailties, and (6) a Cox model with lognormal frailties.

	Adaptive: Poisson(20)			Adaptive: Geometric(0.1)		
Covariate	PM	2.5%	97.5%	PM	2.5%	97.5%
Adult Onset	0.221	−0.256	0.722	0.247	−0.485	0.822
Treatment	−0.619	−1.024	−0.189	−0.622	−1.171	−0.155
Interaction	−0.805	−1.503	−0.172	−0.803	−1.533	−0.086
Frailty Variance	1.285	0.842	1.772	1.126	0.672	1.748

	Penalized: 2^{nd} diff.			Penalized: 2^{nd} deriv.		
Covariate	PM	2.5%	97.5%	PM	2.5%	97.5%
Adult Onset	0.303	−0.170	0.774	0.365	−0.130	0.893
Treatment	−0.549	−0.968	−0.133	−0.515	−0.948	−0.085
Interaction	−0.785	−1.434	−0.157	−0.846	−1.583	−0.162
Frailty Variance	0.986	0.164	2.076	0.938	0.364	1.429

	Cox: gamma frailty			Cox: lognormal frailty		
Covariate	Est	SD	Pval	Est	SD	Pval
Adult Onset	0.397	0.259	0.126	0.399	0.245	0.104
Treatment	−0.506	0.225	0.025	−0.500	0.225	0.027
Interaction	−0.985	0.362	0.006	−0.966	0.361	0.008
Frailty Variance	0.927			0.832		

penalty on the sum of squared second differences between the spline parameters following Section 2.A.2, with hyperparameters tuned to result in smooth hazard and frailty curves. The fourth is similar, but penalizes the integrated squared second derivative of each curve, as described in Section 2.A.3. In addition, two Cox model fits are shown for comparison, with gamma and lognormal frailties respectively. The resulting regression coefficient and frailty variance posterior mean estimates are close, generally falling well each others' posterior credible intervals. This suggests that unless one wishes to capture unusual features of the hazard

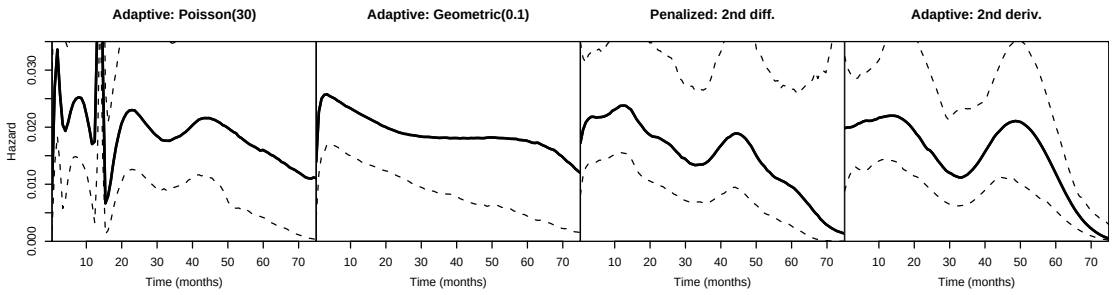

Figure 2.9: Baseline hazard function estimates for the diabetic retinopathy data, under various forms of smoothing: adaptive knot selection with Poisson and Geometric priors on the number of knots, and penalized smoothers with penalties on the sum of squared second differences and integrated squared second derivative, respectively. The baseline is an untreated patient with juvenile-onset diabetes.

or frailty curve, such as the peak shown in Figure 2.8, the choice of smoothing mechanism may not be important in practice. Estimated baseline hazard curves for each of the four fits are shown in Figure 2.9.

For the fit penalized by the sum of squared second differences, we show the hazard, survivor function and frailty density in Figure 2.10. The survivor curve is smoother than the one in Figure 2.8, a result of the control offered by the penalized smoother. There is a strong frailty effect, with the frailty density showing hints of bimodality, possibly suggesting that there may be an additional important binary factor not captured by the covariates. The posterior distributions of the individual frailties also show that the frailty plays a significant role in each patient's overall risk, and additional risk factors beyond treatment and age of onset affect the risk of blindness.

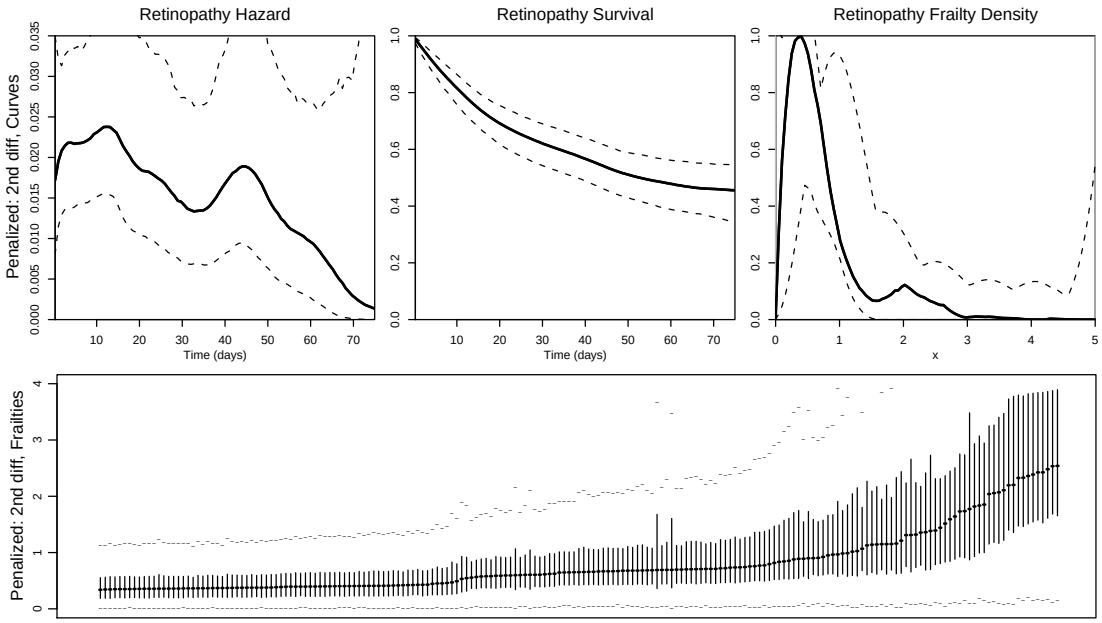

Figure 2.10: Upper panel: Hazard, survival and frailty density estimates, and 95% pointwise posterior intervals for the diabetic retinopathy data. Spline-only specification, with fixed equally spaced knots and smoothness controlled by a penalty on the sum of squared second differences of the spline parameters. Lower panel: Median, interquartile range, and full range of posterior frailty estimates for each of the patients, sorted by posterior median.

2.5 Discussion

The proposed approach permits the analysis of clustered survival data when the underlying frailty distribution is unknown, without being subject to model error. The nonparametric Bayesian approach allows even unusual baseline hazards and frailty distributions to be correctly identified, and, with properly chosen priors, gives accurate posterior means and credible intervals for all parameters involved.

The adaptive knot selection approach allows a simpler model specification than the penalized spline approach. Rather than having to construct an exotic penalty function, with appropriate priors and hyperparameters, the smoothness of the

curve is controlled through the prior on the number of knots. Results show that for data sets that can be well-modeled with a parametric hazard or frailty distribution, the inclusion of a parametric component results in smoother and more accurate fits.

Extensions of the method to stratified data, or data with time-dependent covariates are conceptually simple, but this is not currently supported by the accompanying software.

Unlike existing frequentist methods, the Bayesian approach results in a wealth of information about the joint posterior distribution of all parameters of interest. Posterior estimates of the hazard and survival, and predictions for different risk groups can incorporate the dependence between all parameters, allowing a more thorough understanding of the sources of risk. Furthermore, through deliberate specification of priors, the Bayesian approach allows practitioners to obtain the desired degree of smoothness in the hazard functions and frailty densities, without obscuring important effects.

The method's flexibility comes at the cost of being very computation-intensive. The computational effort involved in evaluating B-splines, computing conditional likelihoods, and calculating penalties over thousands of MCMC iterations is considerable, and fitting a large sample can take several hours on consumer workstations.

Simulation results indicate that the method performs well, especially when the data contain many clusters of reasonable size, a common situation in multicenter clinical studies. In such settings, the gain in flexibility from a fully nonparametric approach may offset the increased computational cost.

APPENDIX

Appendix 2.A Choice of penalty functions

2.A.1 Gaussian penalty

The penalty function may be chosen to yield a Gaussian prior on the parameter set, that is,

$$p_\lambda(\boldsymbol{\theta}_\lambda) = \boldsymbol{\theta}_\lambda^T \boldsymbol{\theta}_\lambda \,,$$

and analogously for $p_u(\boldsymbol{\theta}_u)$. The gradients and Hessians are then

$$\nabla p_\lambda(\boldsymbol{\theta}_\lambda) = 2\boldsymbol{\theta}_\lambda \,, \qquad \nabla^2 p_\lambda(\boldsymbol{\theta}_\lambda) = 2D(\mathbf{1}_{K_u}) \,,$$

and analogously for $p_u(\boldsymbol{\theta}_u)$. This penalty function is recommended when adaptive knot selection is used, since in that case, the smoothness of the curve is controlled through the prior on the number of knots, and does not need to be explicitly penalized.

2.A.2 Penalty on second differences

Let $\boldsymbol{D}$ be a matrix so that $\boldsymbol{Dy}$ computes the second difference in $\boldsymbol{y}$, and let $\boldsymbol{P} = \boldsymbol{D}^T\boldsymbol{D}$. Then, for analogously defined matrices $\boldsymbol{P}_\lambda, \boldsymbol{P}_u$ of the appropriate dimensions, the following functions penalize the second differences in the spline parameters:

$$p_\lambda(\boldsymbol{\theta}_\lambda) = \boldsymbol{\theta}_\lambda^T \boldsymbol{P}_\lambda \boldsymbol{\theta}_\lambda \,, \qquad p_u(\boldsymbol{\theta}_u) = \boldsymbol{\theta}_u^T \boldsymbol{P}_u \boldsymbol{\theta}_u \,,$$

with gradients

$$\nabla p_\lambda(\boldsymbol{\theta}_\lambda) = 2\boldsymbol{P}_\lambda \boldsymbol{\theta}_\lambda \,, \qquad \nabla p_u(\boldsymbol{\theta}_u) = 2\boldsymbol{P}_u \boldsymbol{\theta}_u \,,$$

and hessians

$$\nabla^2 p_\lambda(\boldsymbol{\theta}_\lambda) = 2\boldsymbol{P}_\lambda, \qquad \nabla^2 p_u(\boldsymbol{\theta}_u) = 2\boldsymbol{P}_u \ .$$

While this choice of penalty function is appropriate when the knots are equally spaced, it does not result in smooth behavior otherwise.

2.A.3 Penalty on the second derivative

In order to ensure smoothness even when knots are not equally spaced, we can construct a penalty on the second derivative of the spline. In the case of the baseline hazard spline, that is

$$\begin{aligned} p_\lambda(\boldsymbol{\theta}_\lambda) &= \int_0^\infty \left(\lambda_0^{(2)}(t, \boldsymbol{\theta}_\lambda)\right)^2 dt \\ &= \int_0^\infty \left(\sum_{k=1}^{K_\lambda} B_{\lambda k}^{(2)}(t) \exp(\theta_{\lambda k})\right)^2 dt \\ &= {e^{\boldsymbol{\theta}_\lambda}}^T \boldsymbol{P}_\lambda e^{\boldsymbol{\theta}_\lambda} \ , \end{aligned}$$

where $\boldsymbol{P}_\lambda$ is a matrix whose (j,k) entry is

$$P_{\lambda,jk} = \int_0^\infty B_{\lambda j}^{(2)}(t) B_{\lambda k}^{(2)}(t)\, dt \ . \tag{2.22}$$

This penalty matrix can be computed using a recurrence relation given later in appendix 2.C. The gradient and hessian of the penalty function are then given by

$$\begin{aligned} \nabla p_\lambda(\boldsymbol{\theta}_\lambda) &= 2D(e^{\boldsymbol{\theta}_\lambda})\boldsymbol{P}_\lambda e^{\boldsymbol{\theta}_\lambda} \\ \nabla^2 p_\lambda(\boldsymbol{\theta}_\lambda) &= 2D(e^{\boldsymbol{\theta}_\lambda})\boldsymbol{P}_\lambda D(e^{\boldsymbol{\theta}_\lambda}) + 2D(\boldsymbol{P}_\lambda e^{\boldsymbol{\theta}_\lambda}) D(e^{\boldsymbol{\theta}_\lambda}) \end{aligned}$$

We can construct an analogous penalty matrix for the frailty density, keeping in mind that the frailty density uses normalized splines, that is,

$$p_u(\boldsymbol{\theta}_u) = \frac{{e^{\boldsymbol{\theta}_u}}^T \tilde{\boldsymbol{P}}_u e^{\boldsymbol{\theta}_u}}{(\mathbf{1}_{K_u}^T e^{\boldsymbol{\theta}_u})^2} \ ,$$

where $\tilde{\boldsymbol{P}}_u$ is defined analogously, with the addition of a normalizing factor:

$$\tilde{P}_{u,jk} = \int_0^\infty \tilde{B}^{(2)}_{uj}(t)\tilde{B}^{(2)}_{uk}(t)\,dt$$

Appendix 2.B Choice of parametric components

Both the baseline hazard and frailty density may have optional parametric components. In this section, we present some of the possible choices of distributions, along with appropriate priors, initial values and estimation procedures.

2.B.1 Exponential baseline hazard

The exponential baseline hazard can be parametrized by a constant log-hazard $\boldsymbol{\eta}_\lambda = \eta_\lambda$, so that the hazard function is

$$\lambda_{0p}(t, \boldsymbol{\eta}_\lambda) = \exp(\eta_\lambda)\ , \qquad \Lambda_{0p}(t, \boldsymbol{\eta}_\lambda) = t\exp(\eta_\lambda)\ .$$

A reasonable prior for η_λ is Gaussian with variance σ^2_λ:

$$\log \pi_\lambda(\eta_\lambda|\sigma^2_\lambda) = -\frac{1}{2}\log\sigma^2_\lambda - \frac{\eta^2_\lambda}{2\sigma^2_\lambda}\ ,$$

with an inverse-Gamma prior for σ^2_λ:

$$\log \pi_{\sigma^2_\lambda}(\sigma^2_\lambda|\boldsymbol{\alpha}_{\sigma^2_\lambda}) = -(\alpha_{\sigma^2_\lambda 1}+1)\log\sigma^2_\lambda - \frac{\alpha_{\sigma^2_\lambda 2}}{\sigma^2_\lambda}$$

depending on hyperparameters $\alpha_{\sigma^2_\lambda 1}, \alpha_{\sigma^2_\lambda 2}$ fixed at 0.01.

Candidates for the $k+1$-st iteration η_λ may be generated as $N(\eta^{(h)}_\lambda, \gamma_{\eta_\lambda})$, where γ_{η_λ} is a tuning parameter chosen to make the acceptance probability close to 25%.

2.B.2 Weibull baseline hazard

The Weibull baseline hazard is parametrized by a log-hazard $\eta_{\lambda 1}$ and log scale parameter $\eta_{\lambda 2}$, so the hazard function is

$$\lambda_{0p}(t, \boldsymbol{\eta}_\lambda) = \exp(\eta_{\lambda 1} + \eta_{\lambda 2} + (e^{\eta_{\lambda 2}} - 1)\log t)\,, \qquad \Lambda_{0p}(t, \boldsymbol{\eta}_\lambda) = \exp(\eta_{\lambda 1}) t^{\exp(\eta_{\lambda 2})}$$

Similar to the exponential case, assume the priors for $\boldsymbol{\eta}_\lambda$ are independent Gaussian with variances $\boldsymbol{\sigma}^2_\lambda = (\sigma^2_{\lambda 1}, \sigma^2_{\lambda 2})$:

$$\log \pi_\lambda(\eta_{\lambda i} | \boldsymbol{\sigma}^2_\lambda) = -\sum_i \log \sigma^2_{\lambda i} - \frac{\eta^2_{\lambda i}}{2\sigma^2_{\lambda_i}}\,,$$

with inverse-Gamma priors for $\sigma^2_{\lambda i}$:

$$\log \pi_{\sigma^2 \lambda i}(\sigma^2_{\lambda i} | \boldsymbol{\alpha}_{\sigma^2_\lambda i}) = -(\alpha_{\sigma^2_\lambda i 1} + 1)\log \sigma^2_\lambda - \frac{\alpha_{\sigma^2_\lambda i 2}}{\sigma^2_\lambda}$$

depending on hyperparameters $\alpha_{\sigma^2_\lambda i 1}, \alpha_{\sigma^2_\lambda i 2}$ fixed at 0.01.

Candidates for the $k+1$-st iteration $\boldsymbol{\eta}_\lambda$ may be generated independently as $N(\eta^{(h)}_{\lambda i}, \gamma_{\eta_\lambda i})$, where $\gamma_{\eta_\lambda i}$ are tuning parameters chosen to make the acceptance probability close to 25%. It is possible to simplify the prior structure somewhat by assuming that $\sigma^2_{\lambda 1} = \sigma^2_{\lambda 2} = \sigma^2_\lambda$, and we have found this to be equally effective.

2.B.3 Gamma frailty density

The gamma frailty distribution parametrized by its log-variance η_u results in the following parametric density component:

$$f_p(x, \eta_u) = x^{\exp(-\eta_u)-1} \frac{\exp(-\eta_u)^{\exp(-\eta_u)} e^{-\exp(-\eta_u)x}}{\Gamma(\exp(-\eta_u))}$$

Similar to the Exponential baseline hazard case case, let the the prior for η_u be univariate Gaussian with variance σ^2_u, in which case the hierarchical structure and candidate generation is identical.

2.B.4 Lognormal frailty density

The lognormal frailty density parametrized by a log-variance parameter η_u results in the following parametric density component:

$$f_p(x, \eta_u) = \frac{\exp\left(-\frac{(\log x + \frac{1}{2}\exp(\eta_u))^2}{2\exp(\eta_u)}\right)}{x\sqrt{2\pi\exp(\eta_u)}}$$

This is a lognormal distribution with mean 1 and variance $e^{\exp(\eta_u)} - 1$. Choosing a Gaussian prior for the log-variance allows a parametrization identical to the Gamma case above.

Appendix 2.C Computing integrals over the B-splines

Consider first the B-spline $\boldsymbol{B}_\lambda(t)$ specifying the baseline hazard. Recall that events occur at times T_{ij}, and the number of interior knots is set to

$$N_\lambda = \min\left(\frac{\sum_i J_i}{4}, 35\right) ,$$

and the knots will be positioned at $\xi_{\lambda k} = \min(T_{ij}) + k\Delta_\lambda$, for $k = 0, \ldots, N_\lambda + 1$, where

$$\Delta_\lambda = \frac{\max(T_{ij}) - \min(T_{ij})}{N_\lambda + 1} .$$

In addition, if the spline order $Q_\lambda > 1$, define repeated exterior knots located at the boundaries, so that $\xi_{\lambda(-1)}, \ldots, \xi_{\lambda(-Q_\lambda+1)} = \xi_{\lambda 0}$ and $\xi_{\lambda(N_\lambda+1)}, \ldots, \xi_{\lambda(N_\lambda+Q_\lambda)} = \xi_{\lambda N_\lambda}$. Note that $B_{\lambda k}$ is supported on the range $(\xi_{\lambda(k-Q_\lambda)}, \xi_{\lambda k})$.

For the (unnormalized) B-spline $\boldsymbol{B}_u(x)$ of order Q_u defining the frailty density, the number and placement of knots ξ_{uk} are determined analogously.

In this section we give formulas for the integrals $C_\lambda(T_{ij})$ (the cumulative baseline hazard), E_{uk} (the mean of each normalized spline component), the normalization factor to produce normalized B-splines $\tilde{\boldsymbol{B}}_u(x)$, and the integrals for a penalty over the second derivative.

2.C.1 Cumulative hazard and normalization factor

A formula for the indefinite integral of a B-spline is given in Cox (1982), allowing us to compute

$$C_{\lambda k}(x) = \int_0^x B_{\lambda k}(t)dt = \begin{cases} \frac{\xi_{\lambda k} - \xi_{\lambda(k-Q_\lambda)}}{Q_\lambda} \sum_{k'=k+1}^{k+Q_\lambda} B'_{\lambda k'}(t) & \text{if} \quad \xi_{\lambda(k-Q_\lambda)} \le t < \xi_{\lambda k} \\ \frac{\xi_{\lambda k} - \xi_{\lambda(k-Q_\lambda)}}{Q_\lambda} & \text{if} \quad t \ge \xi_{\lambda k} \\ 0 & \text{otherwise.} \end{cases}$$

where $B'_{\lambda k'}$ are splines of order $Q_\lambda + 1$ defined on the same set of knots. It follows from an analogue of this formula that the normalized B-splines for the frailty density are defined as

$$\tilde{B}_{uk}(x) = \frac{Q_u}{\xi_{uk} - \xi_{u(k-Q_u)}} B_{uk}(x) \ .$$

2.C.2 Moments of a normalized B-spline

In order to compute E_{uk} denote

$$M_{n,q,k} = \int_{-\infty}^{\infty} x^n \tilde{B}_{q,k}(x)\, dx$$

where $\tilde{B}_{q,k}$ is a normalized B-spline of order q with knots $\xi_{-q+1}, \ldots, \xi_{K+q}$. This quantity can be thought of as the n-th moment of a random variable whose density is given by a single normalized B-spline. It is easy to show (using integration by

parts and a recurrence relation for the derivative of a B-spline), that $M_{n,q,k}$ satisfies the recurrence relation

$$M_{n,q,k} = \begin{cases} \frac{q}{\xi_k - \xi_{k-q}} \cdot \frac{1}{n+1} \big[- M_{n+1,q-1,k-1} + M_{n+1,q-1,k} \big] & \text{if } \xi_k - \xi_{k-q} > 0 \\ \xi_k^n & \text{otherwise.} \end{cases}$$

Given this, the terms E_{uk} can be computed as

$$E_{uk} = 1 - M_{1,Q_u,k} \ ,$$

using the knots ξ_{uk} corresponding to the normalized B-splines $\tilde{B}_{uk}$.

2.C.3 Construction of the penalty matrix on the integrated squared second derivative

The penalty matrix in eq. (2.22) can be computed by two recurrence relations. These relations follow from recurrence relations on B-splines and their derivatives, and integration by parts.

Let

$$\begin{aligned} f(q_1, k_1, \ell_1, q_2, k_2, \ell_2) &= \int_{-\infty}^{\infty} B_{q_1,k_1}^{(\ell_1)}(x) B_{q_2,k_2}^{(\ell_2)}(x) dx \\ g(n, q_1, k_1, q_2, k_2) &= \int_{-\infty}^{\infty} x^n B_{q_1,k_1}(x) B_{q_2,k_2}(x) dx \end{aligned}$$

where $B_{q,k}^{(\ell)}$ is the ℓ-th derivative of a spline of order q supported on knots (ξ_{k-q}, ξ_k). Then, the following recurrence relation hold (assume without loss of generality that $\ell_1 \geq \ell_2, q_1 \geq q_2, k_1 \geq k_2$):

$f(q_1, k_1, \ell_1, q_2, k_2, \ell_2) =$

$$\begin{cases} 0 & \text{if } k_1 - q_1 \geq k_2 \\ g(0, q_1, k_1, q_2, k_2) & \text{if } \ell_1 = \ell_2 = 0 \\ \frac{q_1 - 1}{\xi_{k_1-1} - \xi_{k_1-q_1}} f(q_1 - 1, k_1 - 1, \ell_1 - 1, q_2, k_2, \ell_2) & \text{otherwise} \\ \quad - \frac{q_1 - 1}{\xi_{k_1} - \xi_{k_1-q_1+1}} f(q_1 - 1, k_1, \ell_1 - 1, q_2, k_2, \ell_2) & \end{cases}$$

$g(n, q_1, k_1, q_2, k_2) =$

$$\begin{cases} 0 & \text{if } k_1 - q_1 \geq k_2 \\ \frac{1}{n+1} \left(\xi_{k_1}^{n+1} - \xi_{k_1-1}^{n+1} \right) & \text{if } q_1 = q_2 = 1 \\ \frac{1}{\xi_{k_1-1} - \xi_{k_1-q_1}} g(n+1, q_1 - 1, k_1 - 1, q_2, k_2) & \text{otherwise.} \\ \quad - \frac{\xi_{k_1-q_1}}{\xi_{k_1-1} - \xi_{k_1-q_1}} g(n, q_1 - 1, k_1 - 1, q_2, k_2) & \\ \quad + \frac{\xi_{k_1}}{\xi_{k_1} - \xi_{k_1-q_1+1}} g(n, q_1 - 1, k_1, q_2, k_2) & \\ \quad - \frac{1}{\xi_{k_1} - \xi_{k_1-q_1+1}} g(n+1, q_1 - 1, k_1, q_2, k_2) & \end{cases}$$

Each entry in the penalty matrix of eq. (2.22), can be separately computed by these recurrence relations.

Appendix 2.D Gradients and Hessians

This section includes several gradients and Hessians of the loglikelihoods in Section 2.2 needed to generate candidates for the MCMC steps. Many of these parameters vary with the iteration h – this dependence is implicit.

The gradient and Hessian of the loglikelihood for the regression coefficients $\boldsymbol{\beta}$ in eq. (2.15) are given by:

$$\nabla \ell(\boldsymbol{\beta} | \ldots) \;=\; \boldsymbol{Z}^T \left[\boldsymbol{\delta} - D(\boldsymbol{\Lambda}_0) D(e^{\boldsymbol{Z}\boldsymbol{\beta}}) \tilde{\boldsymbol{U}} \right] - \frac{1}{\sigma_\beta^2} \boldsymbol{\beta} \tag{2.23}$$

$$\nabla^2 \ell(\boldsymbol{\beta}|\ldots) \;=\; -\boldsymbol{Z}^T D(\boldsymbol{\Lambda}_0) D(e^{\boldsymbol{Z\beta}}) D(\tilde{\boldsymbol{U}}) \boldsymbol{Z} - \frac{1}{\sigma_\beta^2} \boldsymbol{I}_p \tag{2.24}$$

The gradient and Hessian of the loglikelihood for the baseline hazard spline parameters $\boldsymbol{\theta}_\lambda$ in eq. (2.16) are given by:

$$\begin{aligned}
\nabla \ell(\boldsymbol{\theta}_\lambda|\ldots) &= \phi_\lambda D(e^{\boldsymbol{\theta}_\lambda}) \left[\boldsymbol{B}_\lambda^T D(\boldsymbol{\lambda}_0)^{-1} \boldsymbol{\delta} - \boldsymbol{C}_\lambda^T D(e^{\boldsymbol{Z\beta}}) \tilde{\boldsymbol{U}} \right] - \frac{1}{2\sigma_\lambda^2} \nabla p_\lambda(\boldsymbol{\theta}_\lambda) \\
\nabla^2 \ell(\boldsymbol{\theta}_\lambda|\ldots) &= \phi D(e^{\boldsymbol{\theta}_\lambda}) \left[D\left(\boldsymbol{B}_\lambda^T D(\boldsymbol{\lambda}_0)^{-1} \boldsymbol{\delta} \right) - \phi \boldsymbol{B}_\lambda^T D(\boldsymbol{\delta}) D(\boldsymbol{\lambda}_0)^{-2} \boldsymbol{B}_\lambda D(e^{\boldsymbol{\theta}_\lambda}) \right. \\
&\quad \left. - D\left(\boldsymbol{C}_\lambda^T D(e^{\boldsymbol{Z\beta}}) \tilde{\boldsymbol{U}} \right) \right] - \frac{1}{2\sigma_\lambda^2} \nabla^2 p_\lambda(\boldsymbol{\theta}_\lambda)
\end{aligned}$$

The gradient and Hessian of the loglikelihood for the frailty density spline parameters $\boldsymbol{\theta}_u$ in eq. (2.17)are given by:

$$\begin{aligned}
\nabla \ell(\boldsymbol{\theta}_u|\ldots) &= \phi D(e^{\boldsymbol{\theta}_u}) \tilde{\boldsymbol{B}}_u^T D(f_u)^{-1} \mathbf{1}_m - 2M \boldsymbol{E}_u^T e^{\boldsymbol{\theta}_u} D(\boldsymbol{E}_u) e^{\boldsymbol{\theta}_u} - \frac{1}{2\sigma_u^2} \nabla p_u(\boldsymbol{\theta}_u) \\
\nabla^2 \ell(\boldsymbol{\theta}_u|\ldots) &= \phi D(e^{\boldsymbol{\theta}_u}) \left[D\left(\tilde{\boldsymbol{B}}_u^T D(f_u)^{-1} \mathbf{1}_m \right) - \phi \tilde{\boldsymbol{B}}_u{}^T D(f_u)^{-2} \tilde{\boldsymbol{B}}_u D(e^{\boldsymbol{\theta}_u}) \right] \\
&\quad - \frac{1}{2\sigma_u^2} \nabla^2 p_u(\boldsymbol{\theta}_u)
\end{aligned}$$

BIBLIOGRAPHY

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York.

Aslanidou, H., Dey, D. K., and Sinha, D. (1998). Bayesian analysis of multivariate survival data using monte carlo methods. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **26,** 33–48.

Barker, P. and Henderson, R. (2005). Small sample bias in the gamma frailty model for univariate survival. *Lifetime Data Analysis* **11,**.

Biller, C. (2000). Adaptive bayesian regression splines in semiparametric generalized linear models. *Journal of Computational and Graphical Statistics* **9,** 122–140.

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society. Series A* **148,** 82–117.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34,** 187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Cox, M. (1982). Practical spline approximation. In *Topics in Numerical Analysis*, volume 965 of *Lecture Notes in Mathematics*, pages 79–112. Springer.

de Boor, C. (2001). *A Practical Guide to Splines*, volume 27 of *Applied Mathematical Sciences*. Springer.

Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998). Automatic bayesian curve fitting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60,** 333–350.

DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88,** 1055–1071.

Ghidey, W., Lesaffre, E., and Eilers, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics* **60,** 945–953.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* **82,** 711–732.

Huster, W. J., Brookmeyer, R., and Self, S. G. (1989). Modelling paired survival data with covariates. *Biometrics* **45,** 145–156.

Kalbfleisch, J. D. (1978). Non-parametric bayesian analysis of survival time data. *Journal of the Royal Statistical Society. Series B (Methodological)* **40,** 214–221.

Kim, Y. and Lee, J. (2003). Bayesian analysis of proportional hazard models. *The Annals of Statistics* **31,** 493–511.

Komárek, A. (2006). *Accelerated Failure Time Models for Multivariate Interval-Censored Data with Flexible Distributional Assumptions*. PhD thesis, Katholieke Universiteit Leuven.

Komárek, A. and Lesaffre, E. (2006). Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions. Technical report, Katholieke Universiteit Leuven.

Kottas, A. (2006). Nonparametric bayesian survival analysis using mixtures of weibull distributions. *Journal of Statistical Planning and Inference* **136,** 578–596.

Li, Y., Ryan, L., Bellamy, S., and Satten, G. (2003). Inference on clustered survival data using imputed frailties. *Journal of Computational and Graphical Statistics* **12,** 640–662.

Lindstrom, M. J. (2002). Bayesian estimation of free-knot splines using reversible jumps. *Computational Statistics & Data Analysis* **41,** 255–269.

Muliere, P. and Walker, S. (1997). A bayesian non-parametric approach to survival analysis using polya trees. *Scandinavian Journal of Statistics* **24,** 331–340.

Müller, P. and Quintana, F. A. (2004). Nonparametric bayesian data analysis. *Statist. Sci.* **19,** 95–110.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics* **23,** 182–198.

Nielsen, G. G., Sørensen, T. I. A., Gill, R. D., and Andersen, P. K. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19,** 25–43.

Pan, W. (2001). Using frailties in the accelerated failure time model. *Lifetime Data Analysis* **7,** 55–64.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics* **26,** 183–214.

Pennell, M. L. and Dunson, D. B. (2006). Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics* **62,** 1044—1052.

Ruppert, D., Nettleton, D., and Hwang, J. T. G. (2007). Exploring the information in p-values for the analysis and planning of multiple-test experiments. *Biometrics* **63,** 483–495. 2007.

Sinha, D. (1993). Semiparametric bayesian analysis of multiple event time data. *Journal of the American Statistical Association* **88,** 979–983.

Sinha, D. and Dey, D. (1998). Survival analysis using semiparametric bayesian methods. In Dey, D., editor, *Practical Nonparametric and Semiparametric Bayesian Statistics*, volume 133 of *Lecture Notes in Statistics*, chapter 10, pages 195–211. Springer.

Sinha, D. and Dey, D. K. (1997). Semiparametric bayesian analysis of survival data. *Journal of the American Statistical Association* **92,** 1195–1212.

Smith, M. and Kohn, R. (1996). Nonparametric regression using bayesian variable selection. *Journal of Econometrics* **75,** 317–343.

Staudenmayer, J., Ruppert, D., and Buonaccorsi, J. P. (2008). Density estimation in the presence of heteroskedastic measurement error. Accepted by JASA.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.

Therneau, T. M., Grambsch, P. M., and Pankratz, V. S. (2003). Penalized survival models and frailty. *Journal of Computational and Graphical Statistics* **12,** 156–175.

CHAPTER 3

A NESTED FRAILTY MODEL FOR CLUSTERED BIVARIATE RECURRENT EVENTS

Joint work with Robert L. Strawderman[1]

Recurrent events are frequently encountered in longitudinal biomedical and public health studies. The processes of interest may consist of events considered to be of a single type, such as might occur in a study of bladder tumor recurrences (e.g. Byar, 1980), migratory motor complex periods (e.g. Aalen and Husebye, 1991), or hospitalization rates among renal failure patients (e.g. Schaubel and Cai, 2005). Alternatively, more than one type of event may be encountered, such as in a skin cancer study involving two distinct types of lesions (e.g. Abu-Libdeh et al., 1990) or in a study involving preschool children with asthma where both hospitalizations and physician office visits are tracked (e.g. Cai and Schaubel, 2004). Recurrent episode data, in which subjects may alternate between two states (e.g., symptomatic vs. asymptotic disease states), may also be viewed as a special case of bivariate recurrent event data; see, for example, Cook et al. (1999).

Regression models appropriate for single-type recurrent outcome data have been well-studied in the survival analysis literature. Broadly speaking, important objectives here may include characterizing the relationship between subject-level characteristics and event occurrences, understanding the dynamics of individual event processes, and describing both within- and between-subject variability. Important early work on this problem began with the suggestion of Cox (1972b) to extend the proportional hazards regression model of Cox (1972a) to the case of a modulated renewal process and the subsequent extensions introduced by Prentice

[1]Submitted to *Computational Statistics & Data Analysis*

et al. (1981) and especially Andersen and Gill (1982) for modeling multivariate counting process data. The literature on this topic has grown rapidly over the past 30 years and is now expansive; a contemporary review of existing parametric and semiparametric models is available in Cook and Lawless (2007).

In regression analyses for event processes consisting of multiple types, the analysis objectives are generally similar. However, one may also be interested in characterizing aspects of the relationship between event processes observed on the same subject. The literature on this problem is considerably less extensive than that for single-type recurrent events. For example, conditional on a multivariate multiplicative random effect, Abu-Libdeh et al. (1990) assume that each of type-specific event processes follow an Andersen-Gill model (Andersen and Gill, 1982). In their model, the baseline intensities are parametrically specified, as is the multivariate frailty distribution. Estimation and inference is based on the corresponding marginal likelihood function. Cook et al. (1999) proposes a similar model for the case of bivariate recurrent event processes, the primary differences being the allowances for stratification and a semi-Markov, rather than Markov, specification of the baseline intensities; see also Ng and Cook (1997) for related work. Though analytically tractable, these intensity-based approaches require the correct specification of the frailty distribution and, in the case of Cook et al. (1999), the parametric baseline hazards for each event type.

Semiparametric models for the analysis of multivariate survival data and recurrent event processes have also been considered. With few exceptions, the proposed methods take a marginal perspective, focusing on the estimation of univariate rate and mean functions rather than modeling the full multivariate intensity function, hence dependence between processes. For example, Ng and Cook (1997, 1999)

propose semiparametric estimators for the marginal rate functions of a bivariate point process, assuming the marginal rates each follow a proportional mean model. Efficiency is improved by introducing a working covariance structure derived under the assumption that the processes follow a bivariate mixed nonhomogeneous Poisson process. Cai and Schaubel (2004) consider the related problem of modeling clustered recurrent processes, instead treating the cluster-level association structure as a nuisance parameter.

Xue (1998), extending earlier work of Xue and Brookmeyer (1996), proposes an interesting alternative for analyzing bivariate survival data. Specifically, under a conditionally specified proportional hazards model, Xue (1998) uses a parametric specification of the baseline hazards but avoids the need to specify the frailty distribution. The latter is achieved by making use of the connections between the likelihood under the resulting hazard model and a certain Poisson regression model in order to derive marginal quasilikehood estimators for both the regression and dependence parameters.

In a related paper, Ma et al. (2003) propose an interesting approach for dealing with clustered univariate survival data based on earlier work in Ma's Ph.D. thesis for fitting certain classes of generalized linear mixed models. Specifically, Ma (1999) considers the use of the best linear unbiased predictor (BLUP) of the random effects in fitting Tweedie exponential dispersion models with multiplicative random effects. Optimal estimating equations in the sense of Godambe (1976) are derived for the regression parameters and consistent estimators are derived for the random effect parameters; see Ma and Jørgensen (2007) for details. A Poisson regression model having a multiplicative random effects structure is one example of such a model. Ma et al. (2003) shows how to make use of the connections between

the partial likelihood under the Cox proportional hazards model and Poisson regression models in deriving estimators for the regression parameters, assuming two levels of nested random effects. In an earlier conference paper, Ma et al. (2001) proposes to use this same idea for modeling gap times in a recurrent event setting, in essence proposing an extension suitable for modeling clustered modulated renewal processes. In estimating the frailties by their best linear unbiased predictors (BLUP), the proposed methodology achieves an advantage over existing methods by only requiring one to specify the first two moments of the underlying frailty distribution. Since the models and methodology proposed in Ma et al. (2001) and Ma et al. (2003) are essentially identical, we refer only to the latter in the following unless a distinction is useful.

In this article, we extend the model and methodology introduced in Ma et al. (2003) to the setting of clustered, paired point processes. Specifically, we assume that two event processes are observed on each subject. The processes, each of which follows a modulated renewal process with a possibly stratum-dependent baseline hazard, are assumed to be independent conditionally on a pair of nested correlated frailties. This multivariate, nested dependence structure permits one to model stratified, clustered, paired point processes of dependent recurrent events by specifying the mean and covariance structure of the random effects, thereby avoiding full parametric specification of the frailty distribution. Estimates for regression and frailty dispersion parameters are obtained by introducing an alternative and useful extension of the auxiliary Poisson modeling framework considered in Ma et al. (2003).

This paper is organized as follows: Section 3.1 introduces the relevant notation and model specification. The estimation procedure is summarized in Section 3.2,

with the associated derivations given in Section 3.3, and related computational concerns discussed in Section 3.4. We propose a few natural extensions and modifications to the methodology in Section 3.5. A simulation study is described and summarized in Section 3.6 and serves to demonstrate the excellent performance of the proposed methods. In Section 3.7, we consider the analyses of two different data sets: a study of the effect of selenium supplementation on the risk of developing two types of skin cancer tumors, and a study of the effect of rhDNase on pulmonary exacerbation episodes of cystic fibrosis patients. We conclude the paper with a brief discussion in Section 3.8. Appendix 3.A contains full simulation results on a number of scenarios.

3.1 Notation and Model

Let the observed data consist of recurrent event outcome and covariate information on m independent clusters of $J_i \geq 1$ subjects, $i = 1 \ldots m$. Specifically, it is assumed that subject (i, j) experiences $N_{ij}^{(d)} \geq 0$ recurrent events of type $d \in \{0, 1\}$, occurring at times $0 = S_{ij0}^{(d)} < S_{ij1}^{(d)} < \ldots < S_{ijN_{ij}^{(d)}}^{(d)}$, prior to some censoring time C_{ij}. Denote the recurrent event counting process for each subject as $\{N_{ij}^{(d)}(t), t \geq 0\}$, so that $N_{ij}^{(d)} = N_{ij}^{(d)}(C_{ij})$; ties are not permitted. It is assumed that each subject has available a set of covariates $\{Z_{ij}(t), t \in [0, C_{ij}]\}$ that may depend on time. In addition, we allow for the possibility that subjects are additionally stratified into p levels; we denote the stratum indicator by $\{L_{ij}(t), t \in [0, C_{ij}]\}$, allowing for the possibility that this may too depend on time. The processes $\{Z_{ij}(t), t \in [0, C_{ij}]\}$ and $\{L_{ij}(t), t \in [0, C_{ij}]\}$ are assumed to have left-continuous sample paths and censoring is assumed to be noninformative in the same sense required in Nielsen et al. (1992).

The correlation between subjects within the same cluster, between event processes on the same subject, and between successive event times for a given event type are captured by correlated pairs of nested frailties. More precisely, the cluster-level frailties for each event type are assumed to be positive and independent, with

$$\mathbb{E}\left[U_i^{(d)}\right] = 1\ , \qquad \mathrm{Var}\left(U_i^{(d)}\right) = \sigma^2_{(d)} \tag{3.1}$$

and $\sigma^2_{(d)} \geq 0,\ d = 0, 1$ for $i = 1 \ldots m$. Subject-level frailties are also assumed to be positive and independent conditional on the cluster-level frailties. Specifically, for $i = 1 \ldots m,\ j = 1 \ldots J_i$, it is assumed that

$$\mathbb{E}\left[U_{ij}^{(d)} | U_i^{(0)} = u_i^{(0)}, U_i^{(1)} = u_i^{(1)}\right] = u_i^{(d)} \tag{3.2}$$

$$\mathrm{Var}\left(U_{ij}^{(d)} | U_i^{(0)} = u_i^{(0)}, U_i^{(1)} = u_i^{(1)}\right) = u_i \nu^2_{(d)} \tag{3.3}$$

$$\mathrm{Cov}\left(U_{ij}^{(0)}, U_{ij}^{(1)} | U_i^{(0)} = u_i^{(0)}, U_i^{(1)} = u_i^{(1)}\right) = \theta\ , \tag{3.4}$$

where $\nu^2_{(d)} \geq 0,\ d = 0, 1$ and $\theta \in \mathbb{R}$. The dependence structure induced by (3.1)–(3.4) further implies that the marginal correlation between the subject-level frailties for the two recurrent event types is given by

$$\rho = \mathrm{Cor}(U_{ij}^{(0)}, U_{ij}^{(1)}) = \theta \prod_{d \in \{0,1\}} \left(\sigma^2_{(d)} + \nu^2_{(d)}\right)^{-\frac{1}{2}}\ ,$$

allowing for the possibility of either positive or negative correlation.

Let $U_*^{(*)} = \left(U_*^{(0)}, U_*^{(1)}\right)$, where $U_*^{(d)} = \left(U_i^{(d)}, U_{ij}^{(d)}; j = 1 \ldots J_i, i = 1 \ldots m\right)$ denotes the set of frailties associated with the d^{th} process, $d = 0, 1$. Conditionally upon the full set of frailties $U_*^{(*)}$, the recurrent event counting processes are assumed to form a multivariate counting process with intensities

$$\lambda_{ij}^{(d)}(t) = \lambda_{0L_{ij}(t)}^{(d)} \left(t - S^{(d)}_{ijN_{ij}^{(d)}(t-)}\right) \cdot U_{ij}^{(d)} e^{\beta^{(d)} Z_{ij}(t)}\, I\{t \leq C_{ij}\},$$

for $j = 1 \ldots J_i,\ i = 1 \ldots m$, where $\beta^{(d)}$ are regression coefficients and $\lambda_{0r}^{(d)}(\cdot)$ is the unspecified stratum-specific baseline hazard for stratum r, $r = 1 \ldots p$. This

conditional intensity formulation is evidently semi-Markov (i.e., given $U_*^{(*)}$); related examples of semi-Markov intensity models may be found in Oakes and Cui (1994); Ng and Cook (1997); Cook et al. (1999); Chang and Wang (1999); Duchateau et al. (2003) and Strawderman (2005, 2006).

3.2 Estimation

Maximum likelihood estimation of the regression, hazard and dispersion parameters under the proposed intensity model involves maximizing the corresponding marginal likelihood function. Typically, some variant of the EM algorithm would be used for this purpose, requiring computation of the frailty BUPs. However, this can be a challenging task unless the probability distribution of $U_*^{(*)}$ has been both parametrically specified and exhibits a rather special structure.

Over the next several subsections and similarly to Ma et al. (2003), an Expectation-Maximization-type (EM) algorithm will instead be developed for estimating all model parameters. The "E" step of the algorithm, given the current values of all model parameters, proceeds by approximating the unobserved frailties using BLUPs, derived in a manner similar to Ma et al. (2003) using a certain pair of auxiliary Poisson regression models. The "M" step of the algorithm has two components. First, updated dispersion parameters are computed using bias-corrected Pearson-type estimators derived from frailty BLUPs. Then, conditionally on the set of estimated (or predicted) frailties, the regression and baseline intensity parameter estimates are obtained by maximizing an appropriate conditional likelihood function. The entire iterative estimation procedure is summarized in Figure 3.1.

Broadly speaking, the proposed algorithm mimics the structure of the EM

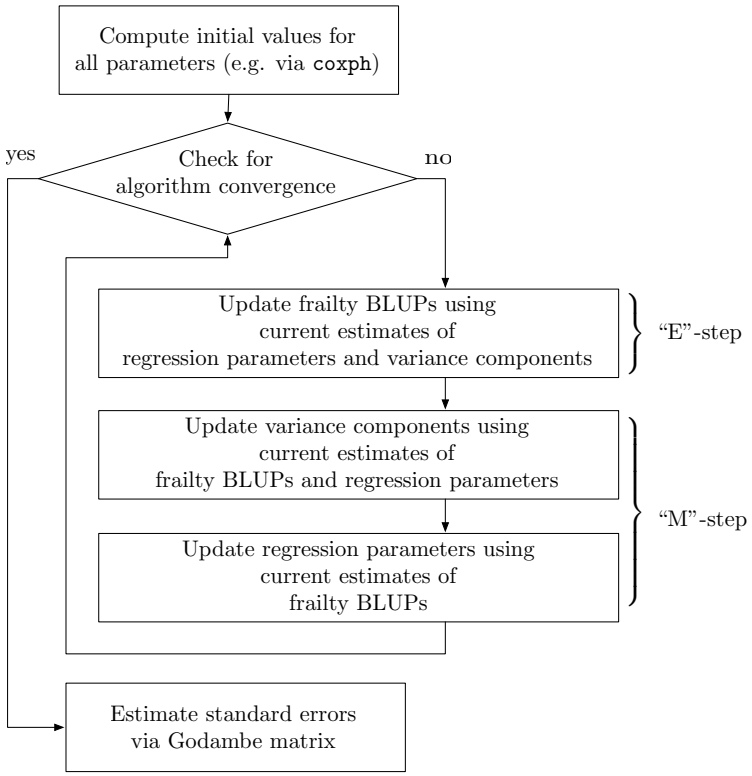

Figure 3.1: An overview of the computational algorithm. Formulas for each of the steps are given in Section 3.2.

algorithm that would be used under a given parametric specification of the frailty distribution. However, it is not a true EM algorithm, for (i) BLUPs shall be used in place of the best unbiased predictors (BUPs) of the random effects that would normally be used in an EM algorithm; and, (ii) Pearson-type estimators will be used in place of the maximum likelihood estimates of the dispersion parameters. The theoretical convergence properties of this algorithm are unknown; however, our practical experience has been that the proposed algorithm is very stable.

The motivation for the proposed "E" and "M" steps requires some further development. In Section 3.2.1, we obtain the conditional likelihood that forms the basis for estimating the regression and baseline intensity parameters. The auxil-

iary Poisson model that we will use for the purpose of deriving the BLUPs of the frailties is introduced in Section 3.2.2. The major components of the estimation scheme summarized in Figure 3.1 are developed in Sections 3.2.3. Specifically, the proposed BLUPs and dispersion parameter estimates are respectively summarized in Sections 3.2.3.1 and 3.2.3.2 and estimation of the regression and baseline intensity parameters is then summarized in 3.2.3.3. The problem of standard error estimation is considered in 3.2.4.

In order to prevent obscuring key ideas, estimation for the proposed model will initially be considered for the case of time-fixed covariates and time-fixed strata; that is, assuming $Z_{ij}(t) = Z_{ij}$ and $L_{ij}(t) = L_{ij}$ for $t \geq 0$, $j = 1 \ldots J_i$, $i = 1 \ldots m$. The generalization to both time-dependent covariates and strata is largely a matter of changing notation; these extensions will be considered briefly in Section 3.2.5.

3.2.1 The Conditional Point Process Likelihood

For $d \in \{0, 1\}$, let $\Lambda_{ij}^{(d)}(t) = \int_0^t \lambda_{ij}^{(d)}(u)du$ and denote the interevent (i.e., gap) times by $T_{ijk}^{(d)} = S_{ijk}^{(d)} - S_{ij,k-1}^{(d)}$, $k = 1 \ldots M_{ij}^{(d)}$. Given $U_*^{(*)}$ and assuming $Z_{ij}(t) = Z_{ij}$ and $L_{ij}(t) = L_{ij}$ for $t \geq 0$, the contribution of subject (i, j) to Jacod's point process likelihood (c.f. Andersen et al., 1993) can be written $\mathcal{L}_{ij}^{(0)}(\lambda_0^{(0)}, \beta^{(0)}|U_*^{(0)}) \times \mathcal{L}_{ij}^{(1)}(\lambda_0^{(1)}, \beta^{(1)}|U_*^{(1)})$, where

$$\mathcal{L}_{ij}^{(d)}(\lambda_0^{(d)}, \beta^{(d)}|U_*^{(d)}) = \prod_{r=1}^{p} \prod_{k=1}^{M_{ij}^{(d)}} \left[\frac{\left(U_{ij}^{(d)} \lambda_{0r}^{(d)}(T_{ijk}^{(d)}) e^{\beta^{(d)} Z_{ij}} \right)^{I(k \leq N_{ij}^{(d)})}}{\exp\left(U_{ij}^{(d)} \Lambda_{0r}^{(d)}(T_{ijk}^{(d)}) e^{\beta^{(d)} Z_{ij}} \right)} \right]^{Y_{rijk}^{(d)}(S_{ijk}^{(d)})} \tag{3.5}$$

and $Y_{rijk}^{(d)}(t) = I(L_{ij} = r) I\left(S_{ij,k-1}^{(d)} < t \leq S_{ijk}^{(d)} \right)$ is an at-risk indicator function that takes on value 1 if subject (i, j) is at risk for the k-th event of type d at time t, while in stratum r. With a single process (i.e., $d = 0$), the likelihood function (3.5)

is equivalent to that considered in Ma et al. (2001); in addition, if $M_{ij}^{(0)} = N_{ij}^{(0)} + 1$ can be at most one for every (i, j) (i.e., at most one event per subject), the resulting likelihood is equivalent to that considered in Ma et al. (2003).

Ma et al. (2003) consider a semiparametric model specification, imposing no assumptions on the baseline functions $\lambda_{r0}(\cdot),\ r = 1 \ldots p$. Such a model specification can lead to an explosion in the total number of parameters, particularly for large sample sizes. The total parameter dimension has a signficant impact on the time required to fit models and may also result in severe numerical instability; see, for example, Ha and Lee (2005). Such problems are compounded by the use of the Newton scoring estimation procedure recommended in Ma et al. (2003), potentially requiring the repeated computation and inversion of high dimensional matrices.

The need to deal with such challenges only increases in the current bivariate setting. A simple method for controlling the dimension of the parameter vector is to employ piecewise constant baseline intensity functions. Suppose that the baseline intensity $\lambda_{0r}^{(d)}(\cdot)$ for a given d and stratum r is finite and piecewise constant on $K_r^{(d)}$ time intervals. Denote the (stratum,process)-specific breakpoints as $0 < a_{r1}^{(d)} < \ldots < a_{rK_r^{(d)}}^{(d)}$; in practice, each interval must contain at least one event, hence selection of these break points will always depend on the observed data. The problem of selecting the number and placement of breakpoints is discussed in Section 3.4. For $t \geq 0$, it is assumed that

$$\lambda_{0r}^{(d)}(t) = \sum_{s=1}^{K_r^{(d)}} \alpha_{rs}^{(d)} I(t \in [a_{r,s-1}^{(d)}, a_{rs}^{(d)})) \tag{3.6}$$

with corresponding cumulative hazards

$$\Lambda_{0r}^{(d)}(t) = \sum_{s=1}^{K_r^{(d)}} \alpha_{rs}^{(d)} A_{rs}^{(d)}(t) \tag{3.7}$$

where $A_{rs}^{(d)}(t) = I\{a_{r,s-1}^{(d)} \leq t\}\left(\min\{t, a_{rs}^{(d)}\} - a_{r,s-1}^{(d)}\right)$. The baseline functions are then fully parameterized through the parameters $\alpha_{rs}^{(d)}$ for each combination of (r, s, d), a number determined by the total number of strata and interval breakpoints. Under a sufficiently fine discretization, the proposed approach reduces to assigning a unique parameter to each unique event time. This is equivalent to specifying semiparametric models for the baseline intensities, thereby generalizing the approach of Ma et al. (2003). As shown in the simulation study, the use of piecewise constant baseline functions has a minimal effect on estimation, provided the discretization level is not too coarse.

Let the discretized at-risk indicators corresponding to the stratum-specific break times be denoted as $Y_{rijks}^{(d)} = Y_{rijk}^{(d)}(a_{rs}^{(d)})$, where $Y_{rijk}^{(d)}(t)$ is defined above. Then, using (3.6) and (3.7), the conditional likelihood (3.5) reduces to

$$\mathcal{L}_{ij}^{(d)}(\alpha^{(d)}, \beta^{(d)}|U_*^{(d)}) = \prod_{r=1}^{p}\prod_{k=1}^{M_{ij}^{(d)}}\prod_{s=1}^{K_r^{(d)}} \frac{\left(U_{ij}^{(d)}\alpha_{rs}^{(d)}e^{\beta^{(d)}Z_{ij}}\right)^{Y_{rijks}^{(d)}\delta_{rijks}^{(d)}}}{\exp\left(Y_{rijks}^{(d)}U_{ij}^{(d)}e^{\beta^{(d)}Z_{ij}}\alpha_{rs}^{(d)}A_{rs}^{(d)}(T_{ijk}^{(d)})\right)}, \tag{3.8}$$

where

$$\delta_{rijks}^{(d)} = I(k \leq N_{ij}^{(d)})I(L_{ij} = r)I(a_{r,s-1}^{(d)} \leq T_{ijk}^{(d)} < a_{rs}^{(d)}). \tag{3.9}$$

The indicator variable $\delta_{rijks}^{(d)} = 1$ may be interpreted as "Subject j in cluster i suffered the k-th event of type d during interval $[a_{r,s-1}^{(d)}, a_{rs}^{(d)})$ while belonging to stratum r." Using (3.8), the full conditional loglikelihood function (i.e., over all subjects) may now be written

$$\ell(\alpha, \beta|U_*^{(*)}) = \sum_{d,r,i,j,k,s} Y_{rijks}^{(d)}\Bigg[\delta_{rijks}^{(d)}\left(\log U_{ij}^{(d)} + \log\alpha_{rs}^{(d)} + \beta^{(d)}Z_{ij}\right) - U_{ij}^{(d)}e^{\beta^{(d)}Z_{ij}}\alpha_{rs}^{(d)}A_{rs}^{(d)}(T_{ijk}^{(d)})\Bigg], \tag{3.10}$$

where α and β respectively denote $([\alpha^{(0)}]^T, [\alpha^{(1)}]^T)^T$ and $([\beta^{(0)}]^T, [\beta^{(1)}]^T)^T$, $U_*^{(*)}$ denotes the entire collection of frailty variables, and the summation appearing out

in front runs over all possible values of (d, r, i, j, k, s), that is,

$$\sum_{d,r,i,j,k,s} x = \sum_{d\in\{0,1\}} \sum_{r=1}^{p} \sum_{i=1}^{m} \sum_{j=1}^{J_i} \sum_{k=1}^{M_{ij}^{(d)}} \sum_{s=1}^{K_r^{(d)}} x \ .$$

3.2.2 An Auxiliary Poisson Model Construction

As indicated in the introduction to Section 3.2, the proposed EM-type algorithm intends to avoid the need to specify the full bivariate frailty distribution by replacing the frailty BUPs with BLUPs derived under an appropriate auxiliary Poisson regression model. One obvious route towards achieving this goal is to propose a direct extension of the methodology in Ma et al. (2003) to the bivariate setting. Specifically, one would begin by noting that the conditional likelihood function

$$\prod_{i=1}^{m} \prod_{j=1}^{J_i} \mathcal{L}_{ij}^{(0)}(\lambda_0^{(0)}, \beta^{(0)} | U_*^{(0)}) \times \mathcal{L}_{ij}^{(1)}(\lambda_0^{(1)}, \beta^{(1)} | U_*^{(1)}).$$

generates a corresponding conditional partial likelihood for β. It is then possible to construct an auxiliary Poisson model, consisting of Poisson regression models specified conditionally on U_* for both $d = 0$ and $d = 1$, that generates a conditional profile likelihood for β equivalent to this conditional partial likelihood. This correspondence is useful because the marginal moment structure of the event indicators under the Poisson model follows directly from the moment structure imposed on the frailties. As a result, one can devise closed-form BLUP expressions without imposing restrictive parametric assumptions on the frailty distribution.

Our initial attempts to implement the above extension to the bivariate setting created numerical problems of the sort described in Section 3.2.1, eventually leading us to consider the discretization (3.6). The use of discretization creates minimal difficulties for the estimation of the regression and hazard parameters;

for example, as shown in Section 3.2.3.3, it is easy to maximize the loglikelihood (3.10) given a set of frailties $U_*^{(*)}$. However, unless the level of discretization is chosen to be sufficiently fine, it proved to be impossible to construct an auxiliary Poisson model that either generates the loglikelihood (3.10) or the corresponding profile loglikelihood for β given in (3.21) below.

Fortunately, an exact correspondence turns out to be unnecessary. The practical importance of the auxiliary Poisson model is limited to its utility in developing BLUP-type approximations to the frailties. We therefore propose to use an auxiliary model that approximates (3.10) under a sufficiently fine level of discretization, restricting the use of this model to the derivation of frailty BLUPs as described below in Section 3.2.3.1. Similarly to Ma et al. (2003), estimates for the dispersion parameters $\sigma^2_{(d)}, \nu^2_{(d)}$ and θ are then derived in Section 3.2.3.2 using Pearson-type estimators, with bias corrections computed under the proposed auxiliary Poisson model.

Specifically, for $d \in \{0,1\}$ and $F^{(d)}_{rijks} = \{Y^{(d)}_{rijks} = 1,\ Z_{ij}, L_{ij} = r\}$, suppose

$$\delta^{(d)}_{rijks} | U^{(*)}_*,\ F^{(d)}_{rijks} \quad \sim \quad \text{Poisson}\left\{U^{(d)}_{ij}\left[1 - \exp\left(-e^{\beta^{(d)} Z_{ij}} \alpha^{(d)}_{rs} h^{(d)}_{rs}\right)\right]\right\}, \qquad (3.11)$$

where $h^{(d)}_{rs} = a^{(d)}_{rs} - a^{(d)}_{r,s-1}$. Assume that these event indicators are mutually independent across all possible combinations of d, r, i, j, k, s indices, conditionally on $U^{(*)}_*$ and all covariate, strata, and at-risk information. The corresponding random effects Poisson loglikelihood function may then be written

$$\ell_A(\alpha, \beta | U^{(*)}_*) = \sum_{d,r,i,j,k,s} Y^{(d)}_{rijks}\Bigg[\delta^{(d)}_{rijks}\left(\log U^{(d)}_{ij} + \log\left[1 - \exp\left(-e^{\beta^{(d)} Z_{ij}} \alpha^{(d)}_{rs} h^{(d)}_{rs}\right)\right]\right) - U^{(d)}_{ij}\left[1 - \exp\left(-e^{\beta^{(d)} Z_{ij}} \alpha^{(d)}_{rs} h^{(d)}_{rs}\right)\right]\Bigg].$$

As $h_{rs}^{(d)} \to 0$, we have $\ell_A(\alpha, \beta | U_*^{(*)}) \approx \widetilde{\ell}_A(\alpha, \beta | U_*^{(*)})$, where

$$\widetilde{\ell}_A(\alpha, \beta | U_*^{(*)}) \propto \sum_{d,r,i,j,k,s} Y_{rijks}^{(d)} \Big[\delta_{rijks}^{(d)} \left(\log U_{ij}^{(d)} + \log \alpha_{rs}^{(d)} + \beta^{(d)} Z_{ij} \right) - U_{ij}^{(d)} e^{\beta^{(d)} Z_{ij}} \alpha_{rs}^{(d)} h_{rs}^{(d)} \Big]. \tag{3.12}$$

Assuming that $h_{rs}^{(d)}$ is small enough to ensure that exactly one event occurs in each interval, it is additionally true that $K_r^{(d)} = \sum_{i,j,k,s} \delta_{rijks}^{(d)}$ and $Y_{rijks}^{(d)} h_{rs}^{(d)} = Y_{rijks}^{(d)} A_{rs}^{(r)}(T_{ijk}^{(d)})$, implying the equivalence of (3.12) and (3.10) under sufficiently fine levels of discretization.

REMARK: To better understand the motivation behind (3.11), suppose there were no censoring. Then, it follows from (3.9) that $\delta_{rijks}^{(d)} = \widetilde{\delta}_{rijks}^{(d)}$, where

$$\widetilde{\delta}_{rijks}^{(d)} = I(L_{ij} = r) I(a_{r,s-1}^{(d)} \leq T_{ijk}^{(d)} < a_{rs}^{(d)}).$$

Under the proposed intensity model and with $F_{rijks}^{(d)}$ as defined earlier,

$$\begin{aligned} \mathbb{E}\left[\widetilde{\delta}_{rijks}^{(d)} \middle| U_*^{(*)},\ F_{rijks}^{(d)} \right] &= \mathbb{P}\left\{ a_{r,s-1}^{(d)} \leq T_{ijk}^{(d)} < a_{rs}^{(d)} \middle| U_{ij}^{(d)}, F_{rij}, T_{ijk}^{(d)} > a_{r,s-1}^{(d)} \right\} \\ &= 1 - \exp\left(- \int_{a_{r,s-1}^{(d)}}^{a_{rs}^{(d)}} \lambda_{0r}^{(d)}(t) U_{ij}^{(d)} e^{\beta^{(d)} Z_{ij}} \right) \\ &= 1 - \exp\left(-U_{ij}^{(d)} e^{\beta^{(d)} Z_{ij}} \alpha_{rs}^{(d)} h_{rs}^{(d)} \right). \end{aligned} \tag{3.13}$$

Observe that (3.13) is a nonlinear function of $U_{ij}^{(d)}$. However, letting $h_{rs}^{(d)} \to 0$ as before, we obtain the approximation

$$\mathbb{E}\left[\widetilde{\delta}_{rijks}^{(d)} \middle| U_*^{(*)}, F_{rijks}^{(d)} \right] \approx U_{ij}^{(d)} \left[1 - \exp\left(-e^{\beta^{(d)} Z_{ij}} \alpha_{rs}^{(d)} h_{rs}^{(d)} \right) \right],$$

completing the motivation for (3.11).

REMARK: The marginal Poisson mean $\mathbb{E}\left[\delta_{rijks}^{(d)} | F_{rijks}^{(d)} \right]$ computed under assumption (3.11) is restricted to lie in the interval $(0, 1)$. Empirically, such a restriction led to substantial improvements in the level of agreement between the

observed and expected numbers of events in comparison to alternative formulations that failed to impose this same restriction. For example, this was observed to be true in comparison with the formulation

$$\delta^{(d)}_{rijks}|U^{(*)}_{*},\ F^{(d)}_{rijks} \quad \sim \quad \text{Poisson}\left\{U^{(d)}_{ij}e^{\beta^{(d)}Z_{ij}}\alpha^{(d)}_{rs}h^{(d)}_{rs}\right\},$$

a choice that corresponds to the Poisson model used in Ma et al. (2003).

3.2.3 Parameter estimation

The main iterative algorithm has already been summarized in Figure 3.1. Each iteration of the algorithm consists of 3 steps, the details of which are now summarized in Sections 3.2.3.1–3.2.3.3.

3.2.3.1 Best linear unbiased predictors for frailties

Under the auxiliary Poisson model introduced in Section 3.2.2, one may construct BLUPs for the $U^{(d)}_i$s and $U^{(d)}_{ij}$s given only the moment assumptions summarized in (3.1)-(3.4). Specifically, given the current set of baseline, regression, and dispersion parameters and extending the results in Ma et al. (2003), these BLUPs may be computed via

$$\begin{aligned}\hat{U}^{(d)}_i &= \mathbb{E}\left[U^{(d)}_i\right] + \text{Cov}\left(U^{(d)}_i, \boldsymbol{\delta}\right)\left(\text{Var}\left(\boldsymbol{\delta}\right)\right)^{-1}\left(\boldsymbol{\delta} - \mathbb{E}\left[\boldsymbol{\delta}\right]\right),\\ \hat{U}^{(d)}_{ij} &= \mathbb{E}\left[U^{(d)}_{ij}\right] + \text{Cov}\left(U^{(d)}_{ij}, \boldsymbol{\delta}\right)\left(\text{Var}\left(\boldsymbol{\delta}\right)\right)^{-1}\left(\boldsymbol{\delta} - \mathbb{E}\left[\boldsymbol{\delta}\right]\right),\end{aligned} \tag{3.14}$$

where $\boldsymbol{\delta}$ denotes the vector of all recurrent event indicators $\delta^{(d)}_{rijks}$ for which $Y^{(d)}_{rijks} = 1$ and the various moments appearing in (3.14) are assumed to be conditional on all covariate and strata information. The results of these calculations

are summmarized below; detailed derivations are provided in Section 3.3.1. Let $\mu^{(d)}_{rijks} = \mathbb{E}\left[\delta^{(d)}_{rijks}|F^{(d)}_{rijks}\right]$ and define

$$\mu^{(d)}_{ij.} = \sum_{r,k,s} Y^{(d)}_{rijks}\mu^{(d)}_{rijks}\,, \qquad \delta^{(d)}_{ij.} = \sum_{r,k,s} Y^{(d)}_{rijks}\delta^{(d)}_{rijks}$$

where, under the auxiliary model (3.11), $\mu^{(d)}_{rijks} = 1 - \exp\left(-e^{\beta^{(d)}Z_{ij}}\alpha^{(d)}_{rs}h^{(d)}_{rs}\right)$. For $d \in \{0,1\}$, the desired BLUPs may then be written as follows:

$$\begin{aligned}\hat{U}^{(d)}_i &= 1 + \sigma^2_{(d)}w_i\Big[\left(1+\sigma^2_{(1-d)}p^{(1-d)}_i\right)\left(P^{(d)}_i - p^{(d)}_i + Q^{(1-d)}_i - q^{(1-d)}_i\right)\\ &\qquad -\sigma^2_{(1-d)}q^{(d)}_i\left(P^{(1-d)}_i - p^{(1-d)}_i + Q^{(d)}_i - q^{(d)}_i\right)\Big] \qquad (3.15)\end{aligned}$$

and

$$\begin{aligned}\hat{U}^{(d)}_{ij} &= \hat{U}^{(d)}_i \cdot w_{ij}\left(1+\nu^2_{(1-d)}\mu^{(1-d)}_{ij.}\right) - \hat{U}^{(1-d)}_i\left(\nu^2_{(d)}q^{(1-d)}_{ij} + \theta p^{(1-d)}_{ij}\right)\\ &\qquad +\nu^2_{(d)}\left(P^{(d)}_{ij} + Q^{(1-d)}_{ij}\right) + \theta\left(P^{(1-d)}_{ij} + Q^{(d)}_{ij}\right), \qquad (3.16)\end{aligned}$$

where

$$\begin{aligned}w_i &= \left(\prod_{d\in\{0,1\}}(1+\sigma^2_{(d)}p^{(d)}_i) - \prod_{d\in\{0,1\}}\sigma^2_{(d)}q^{(d)}_i\right)^{-1},\\ w_{ij} &= \left(\prod_{d\in\{0,1\}}(1+\nu^2_{(d)}\mu^{(d)}_{ij.}) - \prod_{d\in\{0,1\}}\theta\mu^{(d)}_{ij.}\right)^{-1}, \qquad (3.17)\end{aligned}$$

$$\begin{aligned}P^{(d)}_{ij} &= \delta^{(d)}_{ij.}\cdot w_{ij}(1+\nu^2_{(1-d)}\mu^{(1-d)}_{ij.}), \qquad Q^{(d)}_{ij} = -\theta w_{ij}\cdot\delta^{(d)}_{ij.}\mu^{(1-d)}_{ij.}\,,\\ p^{(d)}_{ij} &= \mu^{(d)}_{ij.}\cdot w_{ij}(1+\nu^2_{(1-d)}\mu^{(1-d)}_{ij.})\,, \qquad q^{(d)}_{ij} = -\theta w_{ij}\cdot\mu^{(d)}_{ij.}\mu^{(1-d)}_{ij.}\,, \qquad (3.18)\end{aligned}$$

and

$$P^{(d)}_i = \sum_{j=1}^{J_i} P^{(d)}_{ij}\,, \qquad Q^{(d)}_i = \sum_{j=1}^{J_i} Q^{(d)}_{ij}\,, \qquad p^{(d)}_i = \sum_{j=1}^{J_i} p^{(d)}_{ij}\,, \qquad q^{(d)}_i = \sum_{j=1}^{J_i} q^{(d)}_{ij}.$$

REMARK: For a fixed d and when $\theta = 0$ (i.e., the processes for $d = 0$ and $d = 1$ are assumed uncorrelated), the BLUPs (3.15) and (3.16) are structurally identical to those in Ma et al. (2003, eqns. (11) and (12)).

3.2.3.2 Pearson estimators for frailty dispersion parameters

Assuming $U_*^{(*)}$ were observed, easy computations show

$$\tilde{\tilde{\sigma}}^2_{(d)} \;=\; \frac{1}{m}\sum_{i=1}^{m}(U_i^{(d)}-1)^2\,, \qquad \tilde{\tilde{\nu}}^2_{(d)} \;=\; \frac{1}{m}\sum_{i=1}^{m}\frac{1}{J_i}\sum_{j=1}^{J_i}(U_{ij}^{(d)}-U_i^{(d)})^2\,,$$

and

$$\tilde{\tilde{\theta}} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{J_i}\sum_{j=1}^{J_i}(U_{ij}^{(0)}-U_i^{(0)})(U_{ij}^{(1)}-U_i^{(1)})$$

are respectively unbiased estimators for $\sigma^2_{(d)}$, $\nu^2_{(d)}$, $d \in \{0,1\}$ and θ under the moment assumptions (3.1)-(3.4). Naïve Pearson estimators for the dispersion parameters can thus be obtained directly by respectively replacing $U_i^{(d)}$ and $U_{ij}^{(d)}$ with the BLUPs in (3.15) and (3.16). However, such estimates are generally biased due to the variance shrinkage that occurs as a result of using BLUPs. Using the auxiliary Poisson model, one may derive bias-corrected Pearson estimators; the general form of each estimator is given below:

$$\begin{aligned}
\hat{\sigma}^2_{(d)} &= \frac{1}{m}\sum_{i=1}^{m}\left\{(\hat{U}_i^{(d)}-1)^2 + b_i^{(d)}\right\},\\
\hat{\nu}^2_{(d)} &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{J_i}\sum_{j=1}^{J_i}\left\{(\hat{U}_{ij}^{(d)}-\hat{U}_i^{(d)})^2 + b_{ij}^{(d)}\right\},\\
\hat{\theta} &= \frac{1}{m}\sum_{i=1}^{m}\frac{1}{J_i}\sum_{j=1}^{J_i}\left\{(\hat{U}_{ij}^{(0)}-\hat{U}_i^{(0)})(\hat{U}_{ij}^{(1)}-\hat{U}_i^{(1)}) + b_{ij}^{(*)}\right\},
\end{aligned} \tag{3.19}$$

where the bias corrections $b_i^{(d)}$, $b_{ij}^{(d)}$ and $b_{ij}^{(*)}$ are given by:

$$\begin{aligned}
b_i^{(d)} &= \mathbb{E}\left[(\hat{U}_i^{(d)}-U_i^{(d)})^2\right]\,,\\
b_{ij}^{(d)} &= \mathbb{E}\left[(\hat{U}_i^{(d)}-U_i^{(d)})^2\right] + \mathbb{E}\left[(\hat{U}_{ij}^{(d)}-U_{ij}^{(d)})^2\right] - 2\left\{\sigma^2_{(d)} - \mathrm{Cov}\,(\hat{U}_i^{(d)}, U_{ij}^{(d)})\right\}\,,\\
b_{ij}^{(*)} &= \mathbb{E}\left[(\hat{U}_{ij}^{(0)}-U_{ij}^{(0)})(\hat{U}_{ij}^{(1)}-U_{ij}^{(1)})\right] - \mathrm{Cov}\,(\hat{U}_i^{(0)}, \hat{U}_i^{(1)})\\
&\qquad + \mathrm{Cov}\,(\hat{U}_{ij}^{(0)}, U_{ij}^{(1)}) + \mathrm{Cov}\,(\hat{U}_{ij}^{(1)}, U_{ij}^{(0)})\,.
\end{aligned}$$

A complicated, though closed-form, expression for each estimator is available; these explicit formulas, as well as the lengthy computations required to justify them, are provided in Section 3.3.2.

3.2.3.3 Conditional maximum likelihood estimators for regression coefficients

Given the current predictions of the $U_*^{(*)}$, one can maximize the loglikelihood (3.10) separately with respect to the parameter sets $([\alpha^{(0)}]^T, [\beta^{(0)}]^T)^T$ and $([\alpha^{(1)}]^T, [\beta^{(1)}]^T)^T$. Specfically, given d and fixing $\beta^{(d)}$, the corresponding maximum likelihood estimate for $\alpha_{rs}^{(d)}$ is

$$\hat{\alpha}_{rs}^{(d)} = \frac{\sum_{i,j,k} \delta_{rijks}^{(d)}}{\sum_{i,j,k} Y_{rijks}^{(d)} U_{ij}^{(d)} e^{\beta^{(d)} Z_{ij}} A_{rs}^{(d)}(T_{ijk}^{(d)})} = \frac{n_{rs}^{(d)}}{m_{rs}^{(d)}(\beta^{(d)})}. \tag{3.20}$$

Substituting (3.20) into (3.10) leads to a profile likelihood for $\beta^{(d)}$, or

$$\ell_P(\beta^{(d)}|U_*^{(d)}) \propto \sum_{r,i,j,k,s} Y_{rijks}^{(d)} \left[\delta_{rijks}^{(d)}(\log U_{ij}^{(d)} - \log m_{rs}(\beta^{(d)}) + \beta^{(d)} Z_{ij})\right]. \tag{3.21}$$

The estimate $\hat{\beta}^{(d)}$ can then be obtained as the solution of the score equation

$$\sum_{r,i,j,k,s} Y_{rijks}^{(d)} \left[\delta_{rijks}^{(d)} \left(Z_{ij} - \left(m_{rs}^{(d)}(\beta^{(d)})\right)^{-1} \frac{\partial m_{rs}^{(d)}(\beta^{(d)})}{\partial \beta^{(d)}}\right)\right] = 0, \tag{3.22}$$

where

$$\frac{\partial m_{rs}^{(d)}(\beta^{(d)})}{\partial \beta^{(d)}} = \sum_{i,j,k} Y_{rijks}^{(d)} Z_{ij} U_{ij}^{(d)} e^{\beta^{(d)} Z_{ij}} A_{rs}^{(d)}(T_{ijk}^{(d)}) \ .$$

The estimates for $\alpha^{(d)}$ and $\beta^{(d)}$ derived from (3.20) and (3.22) depend only on linear functions of the $U_{ij}^{(d)}$s. Therefore, computation of the conditional maximum likelihood estimators of $\alpha^{(d)}$ and $\beta^{(d)}$ is possible given the BLUPs derived in Section 3.2.3.1.

3.2.4 Standard errors for the regression coefficients and baseline parameters

Assuming the dispersion parameters are known, an estimate for the asymptotic covariance of the regression coefficients and estimated baseline hazard parameters can be obtained analogously to Ma et al. (2003, Sec. 4.2). Denote the vector of regression parameters and baseline parameters for process (d) by $\gamma^{(d)} = [\log \alpha^{(d)}, \beta^{(d)}]$. Define $x^{(d)}_{rijks} = (0, \ldots, 0, 1, 0, \ldots, Z_{ij}^T)^T$ for $Y^{(d)}_{rijks} = 1$, where the 1 is in the position corresponding to $\alpha^{(d)}_{rs}$. Let $\boldsymbol{X}$ denote a block diagonal matrix whose diagonal blocks contain all vectors $x^{(0)}_{rijks}$ and $x^{(1)}_{rijks}$ respectively. Finally, let $\boldsymbol{\gamma} = [\gamma^{(0)}, \gamma^{(1)}]^T$ and denote by $\hat{\boldsymbol{U}}\boldsymbol{\mu}$ the vector with entries $U^{(d)}_{ij}\mu^{(d)}_{rijks}$ for all i, j, k, s, r, d; note that $\hat{\boldsymbol{U}}\boldsymbol{\mu}$ depends on $\boldsymbol{\gamma}$.

The aforedescribed procedure for obtaining the estimated regression coefficients and baseline hazard parameters can be shown to be equivalent to computing the solution $\widehat{\boldsymbol{\gamma}}$ to $\psi(\boldsymbol{\gamma}) = \mathbf{0}$, where

$$\begin{aligned} \psi(\boldsymbol{\gamma}) &= \boldsymbol{X}^T(\boldsymbol{\delta} - \hat{\boldsymbol{U}}\boldsymbol{\mu}) && (3.23) \\ &= \sum_{i=1}^{m} \boldsymbol{X}_i^T D(\mathbb{E}\left[\boldsymbol{\delta}_i\right])\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}(\boldsymbol{\delta}_i - \boldsymbol{\mu}_i) && (3.24) \end{aligned}$$

The estimating equation (3.23) is obtained by differentiating (3.10) and then replacing all unknown frailties $U^{(d)}_i$ and $U^{(d)}_{ij}$ by the corresponding BLUPs in (3.14). The equivalence between (3.23) and (3.24) relies on arguments similar to those of Ma (1999) and is given in Section 3.3.3.

Suppose that (3.23) has mean zero. Then, under suitable conditions, $\widehat{\boldsymbol{\gamma}}$ is asymptotically normal with a covariance matrix that can be estimated by $[G(\widehat{\boldsymbol{\gamma}})]^{-1}$, where

$$G(\boldsymbol{\gamma}) = S(\boldsymbol{\gamma})V(\boldsymbol{\gamma})^{-1}S(\boldsymbol{\gamma}), \qquad (3.25)$$

$S(\boldsymbol{\gamma}) = \mathbb{E}\left[\frac{\partial \psi(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^T}\right]$ and $V(\boldsymbol{\gamma}) = \mathbb{E}\left[\psi(\boldsymbol{\gamma})^{\otimes 2}\right]$, with $x^{\otimes 2}$ denoting the vector outer product xx^T. The matrix (3.25) is referred to as the Godambe information matrix (Godambe, 1991; Jorgensen and Knudsen, 2004) and plays a role analogous to the Fisher information. The matrices $S(\boldsymbol{\gamma})$ and $V(\boldsymbol{\gamma})$ are respectively referred to as the sensitivity and variability matrices; similarly to Ma et al. (2003, Sec. 4.2), it can be shown that $S(\boldsymbol{\gamma}) = -V(\boldsymbol{\gamma})$, allowing explicit computation of $G(\boldsymbol{\gamma})$.

Specifically, define the notation

$$
\begin{aligned}
S_{i1}^{(d,d)} &= \sum_{j,k,s,r} \mu_{rijks}^{(d)} x_{rijks}^{(d)} (x_{rijks}^{(d)})^T \\
&\quad - \sum_j w_{ij} \left[\nu_{(d)}^2 (1 + \nu_{(1-d)}^2 \mu_{ij.}^{(1-d)}) - \theta^2 \mu_{ij.}^{(1-d)}\right] \left(\sum_{k,s} \mu_{rijks}^{(d)} x_{rijks}^{(d)}\right)^{\otimes 2} \\
S_{i1}^{(d,1-d)} &= -\sum_j \theta w_{ij} \left(\sum_{k,s,r} \mu_{rijks}^{(d)} x_{rijks}^{(d)}\right) \left(\sum_{k,s,r} \mu_{rijks}^{(1-d)} x_{rijks}^{(1-d)}\right)^T \\
S_{i2}^{(0)} &= \begin{bmatrix} \sum_{j,k,s,r} w_{ij} (1 + \nu_{(1)}^2 \mu_{ij.}^{(1)}) \mu_{rijks}^{(0)} x_{rijks}^{(0)} \\ -\sum_{j,k,s,r} \theta w_{ij} \mu_{rijks}^{(1)} x_{rijks}^{(1)} \end{bmatrix} \\
S_{i2}^{(1)} &= \begin{bmatrix} \sum_{j,k,s,r} \theta w_{ij} \mu_{rijks}^{(0)} x_{rijks}^{(0)} \\ -\sum_{j,k,s,r} w_{ij} (1 + \nu_{(0)}^2 \mu_{ij.}^{(0)}) \mu_{rijks}^{(1)} x_{rijks}^{(1)} \end{bmatrix}, \\
H_i &= \begin{bmatrix} S_{i1}^{(0,0)} & S_{i1}^{(0,1)} \\ S_{i1}^{(1,0)} & S_{i1}^{(1,1)} \end{bmatrix},
\end{aligned}
$$

and where all remaining notation not defined here is defined as in Section 3.2.3.1. Then, $G(\boldsymbol{\gamma}) = \sum_{i=1}^m G_i(\boldsymbol{\gamma})$, where

$$
\begin{aligned}
G_i(\boldsymbol{\gamma}) = H_i - w_i \Big\{ &\sigma_{(0)}^2 (1 + \sigma_{(1)}^2 p_i^{(1)}) S_{i2}^{(0)^{\otimes 2}} \\
&- \sigma_{(0)}^2 \sigma_{(1)}^2 q_i^{(0)} S_{i2}^{(0)} \nabla S_{i2}^{(1)} + \sigma_{(1)}^2 (1 + \sigma_{(0)} p_i^{(0)}) S_{i2}^{(1)^{\otimes 2}} \Big\} ,
\end{aligned}
$$

with $x \nabla y = xy^T + yx^T$ denoting the symmetric vector outer product, A detailed derivation of the equivalence $S(\boldsymbol{\gamma}) = -V(\boldsymbol{\gamma})$ and the above explicit expression for $G(\boldsymbol{\gamma})$ may be found in Section 3.3.3.

The simulations of Section 3.6 demonstrate evidence of consistency for all model regression and variance component parameters; plots (not shown) also show evidence of asymptotic normality. The asymptotic covariance matrix $[G(\widehat{\boldsymbol{\gamma}})]^{-1}$ generally provides a reasonable but negatively biased approximation to the empirical standard errors, with improvement being observed with increases in both m and especially $J_1 \ldots J_m$. We suspect that this variance underestimation occurs because the estimating equation (3.23) is not exactly unbiased. More precisely, it possible to show that (3.23) is indeed unbiased under the assumptions that define the auxiliary Poisson model and that the asymptotic variance of the resulting estimator is insensitive to the estimation of the variance components. However, if one merely places data generated under the bivariate point process model of this paper in notational correspondence with the auxiliary Poisson model, neither of these conditions is necessarily guaranteed to hold. As a consequence, and in contrast to Ma (1999, Sec. 5.5.2), the standard errors of the regression coefficients and baseline hazard parameters likely depend on whether the variance components are estimated or assumed known. The estimation procedures considered in Ma et al. (2003) suffer from a similar, if unacknowledged, deficiency.

3.2.5 Time-dependent covariates and strata

The methods presented thus far have assumed that the covariates $Z_{ij}(t) = Z_{ij}$ for $t \geq 0$. However, the proposed methodology really only requires that the covariates are piecewise constant on each discretization interval. Thus, an extension of our methods to the case of time-varying covariates is immediate, provided that the path of each time-varying covariate is assumed left-continuous and piecewise constant. Specifically, denoting $Z^{(d)}_{rijks} = Z_{ij}(a^{(d)}_{r,s-1})$ when $Y^{(d)}_{rijks} = 1$, the likelihood

construction (3.8) can be replicated for such time-dependent covariates as follows:

$$\mathcal{L}_{ij}^{(d)}(\alpha,\beta|U_*^{(d)}) \;=\; \prod_{r=1}^{p}\prod_{k=1}^{M_{ij}^{(d)}}\prod_{s=1}^{K_r^{(d)}} \frac{\left(U_{ij}^{(d)}\alpha_{rs}^{(d)}e^{\beta^{(d)}Z_{rijks}^{(d)}}\right)^{Y_{rijks}^{(d)}\delta_{rijks}^{(d)}}}{\exp\left(Y_{rijks}^{(d)}U_{ij}^{(d)}e^{\beta^{(d)}Z_{rijks}^{(d)}}\alpha_{rs}^{(d)}A_{rs}^{(d)}(T_{ijk}^{(d)})\right)} \;,$$

The auxiliary Poisson model can be constructed analogously, defining

$$\mu_{rijks}^{(d)} = 1-\exp\left(-e^{\beta^{(d)}Z_{rijks}^{(d)}}\alpha_{rs}^{(d)}(a_{rs}^{(d)}-a_{r,s-1}^{(d)})\right) \;.$$

Computations are then carried out exactly as described earlier. Simulation results included in Section 3.A indicate that for large samples with fine discretization, this method performs nearly as well as in the fixed covariate case, though with slightly larger biases. An increased bias can arise in cases where the discretization intervals do not match the times at which the time-dependent covariate changes values. The case of time-dependent stratum membership is handled in exactly the same fashion.

3.3 Derivations

This section gives details on the construction of the estimators summarized in Section 3.2. Section 3.3.1 gives the computation of the frailty BLUPs, and Section 3.3.2 shows how the bias-corrected Pearson estimators are derived. Lastly, Section 3.3.3 details the construction of the sensitivity matrix used to obtain standard error estimates for the regression parameters. The approach extends that used by Ma (1999) to the bivariate case, and accounts for the discretization adjustment of eq. (3.11). Many of the computations in this section are carried out under the auxiliary Poisson model.

3.3.1 Derivation of the frailty best linear unbiased predictors

Recall that the orthodox BLUPs may be computed as

$$\hat{U}_i^{(d)} = \mathbb{E}\left[U_i^{(d)}\right] + \mathrm{Cov}\left(U_i^{(d)}, \boldsymbol{\delta}\right)(\mathrm{Var}\left(\boldsymbol{\delta}\right))^{-1}(\boldsymbol{\delta} - \mathbb{E}\left[\boldsymbol{\delta}\right]) \ ,$$
$$\hat{U}_{ij}^{(d)} = \mathbb{E}\left[U_{ij}^{(d)}\right] + \mathrm{Cov}\left(U_{ij}^{(d)}, \boldsymbol{\delta}\right)(\mathrm{Var}\left(\boldsymbol{\delta}\right))^{-1}(\boldsymbol{\delta} - \mathbb{E}\left[\boldsymbol{\delta}\right]) \ ,$$

where $\boldsymbol{\delta}$ denotes a vector of all recurrent event indicators $\delta_{rijks}^{(d)}$ for which $Y_{rijks}^{(d)} = 1$. Due to independence between the clusters, these expressions can be simplified to

$$\begin{aligned} \hat{U}_i^{(d)} &= \mathbb{E}\left[U_i^{(d)}\right] + \mathrm{Cov}\left(U_i^{(d)}, \boldsymbol{\delta}_i\right)(\mathrm{Var}\left(\boldsymbol{\delta}_i\right))^{-1}(\boldsymbol{\delta}_i - \mathbb{E}\left[\boldsymbol{\delta}_i\right]) \ , \\ \hat{U}_{ij}^{(d)} &= \mathbb{E}\left[U_{ij}^{(d)}\right] + \mathrm{Cov}\left(U_{ij}^{(d)}, \boldsymbol{\delta}_i\right)(\mathrm{Var}\left(\boldsymbol{\delta}_i\right))^{-1}(\boldsymbol{\delta}_i - \mathbb{E}\left[\boldsymbol{\delta}_i\right]) \ , \end{aligned} \tag{3.26}$$

where $\boldsymbol{\delta}_i$ is a vector only of the event indicators $\delta_{rijks}^{(d)}$ in cluster i for which $Y_{rijks}^{(d)} = 1$.

In the following, denote

$$\mu_{rijks}^{(d)} = 1 - \exp\left(-e^{\beta^{(d)} Z_{ij}} \alpha_{rs}^{(d)} h_{rs}^{(d)}\right) \ ,$$

where $h_{rs}^{(d)} = a_{rs}^{(d)} - a_{r,s-1}^{(d)}$, and recall that under the auxiliary Poisson model, $\delta_{rijks}^{(d)} | U_{ij}^{(d)} \sim \mathrm{Poisson}\left(U_{ij}^{(d)} \mu_{rijks}^{(d)}\right)$. The following sections will show how to compute closed-form expressions for the BLUPs. We will first compute the necessary moments, then show how the covariance matrix inverse can be computed, and finally show how the computation simplifies into the expressions of section 3.2.3.1.

3.3.1.1 Frailty and event indicator moments

In order to compute closed-form expressions for the BLUPs of (3.26), it is necessary to compute the form of the covariance matrices $\mathrm{Cov}\left(U_{ij}^{(d)}, \boldsymbol{\delta}_i\right)$ and $\mathrm{Var}\left(\boldsymbol{\delta}_i\right)$ under

the auxiliary Poisson model. Denote by $\mathbf{1}_{(a,b)}$ the Kronecker delta, taking on value 1 if $a = b$ and 0 otherwise. Then,

$$
\begin{aligned}
\mathbb{E}\left[U_i^{(d)}\right] &= 1 \\
\mathbb{E}\left[U_{ij}^{(d)}\right] &= \mathbb{E}\left[\mathbb{E}\left[U_{ij}^{(d)}\middle|U_i^{(d)}\right]\right] = \mathbb{E}\left[U_i^{(d)}\right] = 1 \\
\mathrm{Cov}\left(U_a^{(d)}, U_{ij}^{(d)}\right) &= \mathrm{Cov}\left(\mathbb{E}\left[U_{ij}^{(d)}\middle|U_i\right], \mathbb{E}\left[U_a^{(d)}\middle|U_a\right]\right) = \mathbf{1}_{(a,i)}\sigma_{(d)}^2 \\
\mathrm{Cov}\left(U_{ab}^{(d)}, U_{ij}^{(d)}\right) &= \mathbf{1}_{(a,i)}\mathrm{Cov}\left(U_{ib}^{(d)}, U_{ij}^{(d)}\right) \\
&= \begin{cases} \mathbb{E}\left[(U_i^{(d)})^2\right] - 1 = \sigma_{(d)}^2 & \text{if } i = a, j \neq b \\ \mathrm{Var}\left(U_{ij}^{(d)}\right) = \sigma_{(d)}^2 + \nu_{(d)}^2 & \text{if } i = a, j = b \\ 0 & \text{otherwise} \end{cases} \\
&= \mathbf{1}_{(a,i)}(\sigma_{(d)}^2 + \mathbf{1}_{(b,j)}\nu_{(d)}^2) \\
\mathbb{E}\left[\delta_{rijks}^{(d)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\delta_{rijks}^{(d)}\middle|U_{ij}\right]\right] = \mathbb{E}\left[U_{ij}^{(d)}\right]\mu_{rijks}^{(d)} = \mu_{rijks}^{(d)} \\
\mathrm{Cov}\left(U_{ab}^{(d)}, U_{ij}^{(1-d)}\right) &= \mathbf{1}_{(a,i)}\mathrm{Cov}\left(U_{ib}^{(d)}, U_{ij}^{(1-d)}\right) = \mathbf{1}_{(a,i)}\mathbf{1}_{(b,j)}\theta \\
\mathrm{Cov}\left(U_a^{(d)}, \delta_{rijks}^{(d)}\right) &= \mathrm{Cov}\left(U_a^{(d)}, U_{ij}^{(d)}\right)\mu_{rijks}^{(d)} = \mathbf{1}_{(a,i)}\sigma_{(d)}^2\mu_{rijks}^{(d)} \\
\mathrm{Cov}\left(U_{ab}^{(d)}, \delta_{rijks}^{(d)}\right) &= \mathrm{Cov}\left(U_{ab}^{(d)}, U_{ij}^{(d)}\right)\mu_{rijks}^{(d)} = \mathbf{1}_{(a,i)}(\sigma_{(d)}^2 + \mathbf{1}_{(b,j)}\nu_{(d)}^2)\mu_{rijks}^{(d)} \\
\mathrm{Cov}\left(U_a^{(d)}, \delta_{rijks}^{(1-d)}\right) &= \mathrm{Cov}\left(U_a^{(d)}, U_{ij}^{(1-d)}\right)\mu_{rijks}^{(1-d)} = 0 \\
\mathrm{Cov}\left(U_{ab}^{(d)}, \delta_{rijks}^{(1-d)}\right) &= \mathrm{Cov}\left(U_{ab}^{(d)}, U_{ij}^{(1-d)}\right)\mu_{rijks}^{(1-d)} = \mathbf{1}_{(a,i)}\mathbf{1}_{(b,j)}\theta\mu_{rijks}^{(1-d)} \\
\mathrm{Cov}\left(\delta_{qabce}^{(d)}, \delta_{rijks}^{(d)}\right) &= \mathbb{E}\left[\mathrm{Cov}\left(\delta_{qabce}^{(d)}, \delta_{rijks}^{(d)}\middle|U_{ij}^{(d)}, U_{ab}^{(d)}\right)\right] \\
&\qquad + \mathrm{Cov}\left(\mathbb{E}\left[\delta_{qabce}^{(d)}\middle|U_{ab}^{(d)}\right], \mathbb{E}\left[\delta_{rijks}^{(d)}\middle|U_{ij}^{(d)}\right]\right) \\
&= \mathbf{1}_{(qabce,rijks)}\mathbb{E}\left[U_{ij}^{(d)}\mu_{rijks}^{(d)}\right] + \mathrm{Cov}\left(U_{ab}^{(d)}\mu_{qabce}^{(d)}, U_{ij}^{(d)}\mu_{rijks}^{(d)}\right) \\
&= \mathbf{1}_{(qabce,rijks)}\mu_{rijks}^{(d)} + \mathbf{1}_{(a,i)}(\sigma_{(d)}^2 + \mathbf{1}_{(b,j)}\nu_{(d)}^2)\mu_{qabce}^{(d)}\mu_{rijks}^{(d)} \\
\mathrm{Cov}\left(\delta_{qabce}^{(d)}, \delta_{rijks}^{(1-d)}\right) &= 0 + \mathrm{Cov}\left(U_{ab}^{(d)}\mu_{qabce}^{(d)}, U_{ij}^{(1-d)}\mu_{rijks}^{(1-d)}\right) \\
&= \mathbf{1}_{(a,i)}\mathbf{1}_{(b,j)}\theta\mu_{qabce}^{(d)}\mu_{rijks}^{(1-d)}
\end{aligned}
$$

All other covariances are zero. Thus, the covariance matrices relevant to expression (3.26) can be expressed as:

$$\begin{aligned} \text{Cov}\left(U_i^{(d)}, \boldsymbol{\delta}_i^{(e)}\right) &= \mathbf{1}_{(d,e)}\sigma^2_{(d)}\boldsymbol{\mu}_i^{(d)} \\ \text{Cov}\left(U_{ij}^{(d)}, \boldsymbol{\delta}_i^{(e)}\right) &= \mathbf{1}_{(d,e)}\sigma^2_{(d)}\boldsymbol{\mu}_i^{(d)} + \nu^2_{(d)}\boldsymbol{f}_{ij}^{(d)} \end{aligned} \tag{3.27}$$

where

$$\boldsymbol{f}_{ij}^{(d)} = \left[\mathbf{0}^T_{\boldsymbol{\mu}_{i1}^{(d)}}, \ldots, \mathbf{0}^T_{\boldsymbol{\mu}_{i(j-1)}^{(d)}}, \boldsymbol{\mu}^{(d)T}_{ij}, \mathbf{0}^T_{\boldsymbol{\mu}_{i(j+1)}^{(d)}}, \ldots, \mathbf{0}^T_{\boldsymbol{\mu}_{iJ_i}}\right]^T$$

and $\mathbf{0}_{\boldsymbol{x}}$ denotes a vector of zeros of the same length as $\boldsymbol{x}$. That is, $\boldsymbol{f}_{ij}^{(d)}$ is a vector of zeros of the same length as $\boldsymbol{\mu}_i^{(d)}$, except for $\boldsymbol{\mu}_{ij}^{(d)}$ in the correct position. The variance of event indicators has a block form:

$$\begin{aligned} V_i &= \text{Var}\left(\boldsymbol{\delta}_i\right) = \begin{bmatrix} V_i^{(0,0)} & V_i^{(0,1)} \\ V_i^{(1,0)} & V_i^{(1,1)} \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2_{(0)}\boldsymbol{\mu}_i^{(0)}\boldsymbol{\mu}_i^{(0)T} & 0 \\ 0 & \sigma^2_{(1)}\boldsymbol{\mu}_i^{(1)}\boldsymbol{\mu}_i^{(1)T} \end{bmatrix} + \underbrace{\begin{bmatrix} \tilde{V}_i^{(0,0)} & \tilde{V}_i^{(0,1)} \\ \tilde{V}_i^{(1,0)} & \tilde{V}_i^{(1,1)} \end{bmatrix}}_{\tilde{V}_i}. \end{aligned} \tag{3.28}$$

where

$$\begin{aligned} \tilde{V}_i^{(d,d)} &= \nu^2_{(d)} \begin{bmatrix} \boldsymbol{\mu}_{i1}^{(d)}\boldsymbol{\mu}_{i1}^{(d)T} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \boldsymbol{\mu}_{iJ_i}^{(d)}\boldsymbol{\mu}_{iJ_i}^{(d)T} \end{bmatrix} + D(\boldsymbol{\mu}_i^{(d)}) \\ \tilde{V}_i^{(d,1-d)} &= \theta \begin{bmatrix} \boldsymbol{\mu}_{i1}^{(d)}\boldsymbol{\mu}_{i1}^{(1-d)T} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \boldsymbol{\mu}_{iJ_i}^{(d)}\boldsymbol{\mu}_{iJ_i}^{(1-d)T} \end{bmatrix} \end{aligned} \tag{3.29}$$

where $D(\boldsymbol{x})$ is a diagonal matrix with $\boldsymbol{x}$ on the diagonal.

3.3.1.2 Inverting the covariance matrix

For simplicity, this section will use generic notation that avoids unnecessary subscripts. The results of this section will be applied to the matrices of Section 3.3.1.1 in the later subsections. Suppose we wish to compute the inverse W of a symmetric matrix V of the form

$$V = \begin{bmatrix} axx^T & 0 \\ 0 & byy^T \end{bmatrix} + \underbrace{\begin{bmatrix} V_{00} & V_{01} \\ V_{01}^T & V_{11} \end{bmatrix}}_{\tilde{V}},$$

and suppose that the inverse of $\tilde{V}$ is known and can be written as

$$\tilde{W} = \tilde{V}^{-1} = \begin{bmatrix} A & B \\ B^T & D \end{bmatrix},$$

then denoting $\tilde{x} = [x, \mathbf{0}_y]^T$ and $\tilde{y} = [\mathbf{0}_x, y]^T$ and writing V as

$$V = a\tilde{x}\tilde{x}^T + \underbrace{b\tilde{y}\tilde{y}^T + \tilde{V}}_{\bar{V}}$$

allows the inverse to be computed by two applications of the Sherman-Morrison formula. First, $\bar{V}^{-1}$ is given by

$$\bar{V}^{-1} = \tilde{W} - \frac{b\tilde{W}\tilde{y}\tilde{y}^T\tilde{W}}{1 + b\tilde{y}^T\tilde{W}\tilde{y}} = \tilde{W} - \frac{b\begin{bmatrix} By \\ Dy \end{bmatrix}^{\otimes 2}}{1 + by^TDy}. \tag{3.30}$$

and $W = V^{-1}$ is given by

$$W = \bar{V}^{-1} - \frac{a\bar{V}^{-1}\tilde{x}\tilde{x}^T\tilde{V}^{-1}}{1 + a\tilde{x}^T\tilde{V}^{-1}\tilde{x}} \tag{3.31}$$

In order to simplify equation (3.31), note that

$$\bar{V}^{-1}\tilde{x} = \tilde{W}\tilde{x} - \frac{b\begin{bmatrix} By \\ Dy \end{bmatrix}^{\otimes 2}\tilde{x}}{1 + by^TDy} = \begin{bmatrix} Ax \\ B^Tx \end{bmatrix} - \frac{bx^TBy}{1 + by^TDy}\begin{bmatrix} By \\ Dy \end{bmatrix}. \tag{3.32}$$

Substituting equations (3.30) and (3.32) into (3.31) and simplifying gives an expression for W:

$$
\begin{aligned}
W &= \tilde{W} - \frac{b\begin{bmatrix} By \\ Dy \end{bmatrix}^{\otimes 2}}{1+by^TDy} - \frac{a\left(\begin{bmatrix} Ax \\ B^Tx \end{bmatrix} - \frac{bx^TBy}{1+by^TDy}\begin{bmatrix} By \\ Dy \end{bmatrix}\right)^{\otimes 2}}{1+ax^TAx - \frac{ab(x^TBy)^2}{1+by^TDy}} \\
&= \tilde{W} - \frac{b\begin{bmatrix} By \\ Dy \end{bmatrix}^{\otimes 2}}{1+by^TDy} \\
&\quad - \frac{a(1+by^TDy)\left(\begin{bmatrix} Ax \\ B^Tx \end{bmatrix}^{\otimes 2} - \frac{bx^TBy}{1+by^TDy}\left(\begin{bmatrix} Ax \\ B^Tx \end{bmatrix} \nabla \begin{bmatrix} By \\ Dy \end{bmatrix}\right) + \frac{b^2(x^TBy)^2}{(1+by^TDy)^2}\begin{bmatrix} By \\ Dy \end{bmatrix}^{\otimes 2}\right)}{(1+ax^TAx)(1+by^TDy) - ab(x^TBy)^2} \\
&= \tilde{W} - \frac{a(1+by^TDy)\begin{bmatrix} Ax \\ B^Tx \end{bmatrix}^{\otimes 2} - abx^TBy\left(\begin{bmatrix} Ax \\ B^Tx \end{bmatrix} \nabla \begin{bmatrix} By \\ Dy \end{bmatrix}\right) + b(1+ax^TAx)\begin{bmatrix} By \\ Dy \end{bmatrix}^{\otimes 2}}{(1+ax^TAx)(1+by^TDy) - ab(x^TBy)^2}
\end{aligned}
\tag{3.33}
$$

Note in particular that denoting $w = ((1+ax^TAx)(1+by^TDy) - ab(x^TBy)^2)^{-1}$ allows the easy computation of

$$
\begin{aligned}
\begin{bmatrix} x \\ 0 \end{bmatrix}^T W \begin{bmatrix} z_1 \\ 0 \end{bmatrix} &= x^TAz_1 - w\Big[a(1+by^TDy)\, x^TAx\, x^TAz_1 \\
&\qquad -ab\, x^TBy(x^TAx\, y^TB^Tz_1 + x^TBy\, x^TAz_1) \\
&\qquad +b(1+ax^TAx)x^TBy\, y^TB^Tz_1\Big] \\
&= w\Big[(1+by^TDy)x^TAz_1 - bx^TBy\, y^TB^Tz_1\Big] \qquad (3.34) \\
\begin{bmatrix} x \\ 0 \end{bmatrix}^T W \begin{bmatrix} 0 \\ z_2 \end{bmatrix} &= x^TB^Tz_2 - w\Big[a(1+by^TDy)x^TAx\, x^TBz_2 \\
&\qquad -abx^TBy(x^TAx\, y^TDz_2 + x^TBy\, x^TB^Tz_2) \\
&\qquad +b(1+ax^TAx)x^TBy\, x^TDz_2 \\
&= w\Big[(1+by^TDy)x^TB^Tz_2 - bx^TBy\, y^TDz_2\Big]\,. \qquad (3.35)
\end{aligned}
$$

The sum of eqns (3.34) and (3.35) leads naturally to the cluster-level BLUPs in (3.15) once the appropriate values are substituted in the next section.

3.3.1.3 Computation of cluster-level frailty BLUPs

For purposes of computing the cluster-level BLUPs, denote by $\tilde{W}_i^{(d,e)}$ the (d,e)-quarter-block of the inverse of the matrix $\tilde{V}_i$ in (3.28). Then the inverse of the covariance matrix W_i takes the form of (3.33) for $A = \tilde{W}_i^{(0,0)}$, $B = \tilde{W}_i^{(0,1)}$, $D = \tilde{W}_i^{(1,1)}$, $x = \boldsymbol{\mu}_i^{(0)}$, $y = \boldsymbol{\mu}_i^{(1)}$, $a = \sigma^2_{(0)}$ and $b = \sigma^2_{(1)}$. To compute the cluster BLUPs for $d = 0$, using the covariances from (3.27) gives:

$$\begin{aligned}
\hat{U}_i^{(0)} &= 1 + \sigma^2_{(0)} \begin{bmatrix} \boldsymbol{\mu}_i^{(0)} \\ 0 \end{bmatrix}^T W_i \begin{bmatrix} \boldsymbol{\delta}_i^{(0)} - \boldsymbol{\mu}_i^{(0)} \\ \boldsymbol{\delta}_i^{(1)} - \boldsymbol{\mu}_i^{(1)} \end{bmatrix} \qquad (3.36) \\
&= 1 + \sigma^2_{(0)} w_i \Big[\left(1 + \sigma^2_{(1)} \boldsymbol{\mu}_i^{(1)} \tilde{W}_i^{(1,1)} \boldsymbol{\mu}_i^{(1)}\right) \\
&\quad \cdot \left(\boldsymbol{\mu}_i^{(0)} \tilde{W}_i^{(0,0)} (\boldsymbol{\delta}_i^{(0)} - \boldsymbol{\mu}_i^{(0)}) + \boldsymbol{\mu}_i^{(0)} \tilde{W}_i^{(0,1)} (\boldsymbol{\delta}_i^{(1)} - \boldsymbol{\mu}_i^{(1)})\right) \\
&\quad - \sigma^2_{(1)} \boldsymbol{\mu}_i^{(0)} \tilde{W}_i^{(0,1)} \boldsymbol{\mu}_i^{(1)} \left(\boldsymbol{\mu}_i^{(1)} \tilde{W}_i^{(1,0)} (\boldsymbol{\delta}_i^{(0)} - \boldsymbol{\mu}_i^{(0)}) + \boldsymbol{\mu}_i^{(1)} \tilde{W}_i^{(1,1)} (\boldsymbol{\delta}_i^{(1)} - \boldsymbol{\mu}_i^{(1)})\right) ,
\end{aligned}$$

which is the same as (3.15) for $p_i^{(d)} = \boldsymbol{\mu}_i^{(d)} \tilde{W}_i^{(d,d)} \boldsymbol{\mu}_i^{(d)}$, $P_i^{(d)} = \boldsymbol{\mu}_i^{(d)} \tilde{W}_i^{(d,d)} \boldsymbol{\delta}_i^{(d)}$, $q_i^{(d)} = \boldsymbol{\mu}_i^{(1-d)} \tilde{W}_i^{(1-d,d)} \boldsymbol{\mu}_i^{(d)}$ and $Q_i^{(d)} = \boldsymbol{\mu}_i^{(1-d)} \tilde{W}_i^{(1-d,d)} \boldsymbol{\delta}_i^{(d)}$. The general expression given in (3.15) follows from symmetry. The values of $p_i^{(d)}, P_i^{(d)}, q_i^{(d)}$ and $Q_i^{(d)}$ will be shown to match (3.18) in Section 3.3.1.5.

3.3.1.4 Computation of subject-level frailty BLUPs

Computations analogous to those in (3.34) can be used to show that

$$\begin{aligned}
\begin{bmatrix} r \\ 0 \end{bmatrix}^T W \begin{bmatrix} z_1 - x \\ 0 \end{bmatrix} &= -rAx \left(1 + a \begin{bmatrix} x \\ 0 \end{bmatrix}^T W \begin{bmatrix} z_1 - x \\ 0 \end{bmatrix}\right) \\
&\quad - rBy \left(b \begin{bmatrix} 0 \\ y \end{bmatrix}^T W \begin{bmatrix} z_1 - x \\ 0 \end{bmatrix}\right) + rAz \\
\begin{bmatrix} r \\ 0 \end{bmatrix}^T W \begin{bmatrix} 0 \\ z_2 - y \end{bmatrix} &= -rAx \left(a \begin{bmatrix} x \\ 0 \end{bmatrix}^T W \begin{bmatrix} 0 \\ z_2 - y \end{bmatrix}\right) \\
&\quad - rBy \left(1 + b \begin{bmatrix} 0 \\ y \end{bmatrix}^T W \begin{bmatrix} 0 \\ z_2 - y \end{bmatrix}\right) + rBz_2
\end{aligned}$$

Therefore, noting that by (3.27),

$$\text{Cov}\left(U_{ij}^{(0)}, \boldsymbol{\delta}_i\right) = \begin{bmatrix} \sigma_{(0)}^2 \boldsymbol{\mu}_i^{(0)} + \nu_{(0)}^2 \boldsymbol{f}_{ij}^{(0)} \\ \theta \boldsymbol{f}_{ij}^{(1)} \end{bmatrix},$$

applying the symmetries and simplifying gives:

$$\begin{aligned} \hat{U}_{ij}^{(0)} &= 1 + \hat{U}_i^{(0)} - \left(\nu_{(0)}^2 \boldsymbol{f}_{ij}^{(0)} \tilde{W}_i^{(0,0)} \boldsymbol{\mu}_i^{(0)} + \theta \boldsymbol{f}_{ij}^{(1)} \tilde{W}_i^{(1,0)} \boldsymbol{\mu}_i^{(0)}\right) \hat{U}_i^{(0)} \\ &\quad - \left(\nu_{(0)}^2 \boldsymbol{f}_{ij}^{(0)} \tilde{W}_i^{(0,1)} \boldsymbol{\mu}_i^{(1)} + \theta \boldsymbol{f}_{ij}^{(1)} \tilde{W}_i^{(1,1)} \boldsymbol{\mu}_i^{(1)}\right) U_i^{(1)} \\ &\quad + \nu_{(0)}^2 \left(\boldsymbol{f}_{ij}^{(0)} \tilde{W}_i^{(0,0)} \boldsymbol{\delta}_i^{(0)} + \boldsymbol{f}_{ij}^{(0)} \tilde{W}_i^{(0,1)} \boldsymbol{\delta}_i^{(1)}\right) \\ &\quad + \theta \left(\boldsymbol{f}_{ij}^{(1)} \tilde{W}_i^{(1,0)} \boldsymbol{\delta}_i^{(0)} + \boldsymbol{f}_{ij}^{(1)} \tilde{W}_i^{(1,1)} \boldsymbol{\delta}_i^{(1)}\right), \end{aligned}$$

which matches the expression in (3.16) for $p_{ij}^{(d)} = \boldsymbol{f}_{ij}^{(d)} \tilde{W}_i^{(d,d)} \boldsymbol{\mu}_i^{(d)}$, $P_{ij}^{(d)} = \boldsymbol{f}_{ij}^{(d)} \tilde{W}_i^{(d,d)} \boldsymbol{\delta}_i^{(d)}$, $q_{ij}^{(d)} = \boldsymbol{f}_{ij}^{(1-d)} \tilde{W}_i^{(1-d,d)} \boldsymbol{\mu}_i^{(d)}$ and $Q_{ij}^{(d)} = \boldsymbol{f}_{ij}^{(1-d)} \tilde{W}_i^{(1-d,d)} \boldsymbol{\delta}_i^{(d)}$.

3.3.1.5 Completing the computation

Previous subsections showed that the forms of the cluster- and subject-level frailty BLUPs match those given in Section 3.2.3.1, however, it remains to be shown that the given expressions for $p_{ij}^{(d)}, q_{ij}^{(d)}$, etc. are correct. To that purpose, this section will give the form of $\tilde{W}_i^{(d,d)}$ and $\tilde{W}_i^{(d,1-d)}$ and show how the expressions are computed.

The matrices $\tilde{V}_i^{(d,e)}$ have block-diagonal forms as given in (3.29), and thus their inverses $\tilde{W}_i^{(d,e)}$ are also block-diagonal. Let

$$\begin{aligned} \tilde{V}_{ij}^{(d,d)} &= \nu_{(d)}^2 \boldsymbol{\mu}_{ij}^{(d)} \boldsymbol{\mu}_{ij}^{(d)^T} + D(\boldsymbol{\mu}_{ij}^{(d)}) \\ \tilde{V}_{ij}^{(d,1-d)} &= \theta \boldsymbol{\mu}_{ij}^{(d)} \boldsymbol{\mu}_{ij}^{(1-d)^T} \end{aligned}$$

be the j-th blocks of the matrices $\tilde{V}_i^{(d,d)}$ and $\tilde{V}_i^{(d,1-d)}$ respectively. Note that by the Sherman-Morrison formula,

$$\left(\tilde{V}_{ij}^{(d,d)}\right)^{-1} = D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(d)}}\right) - \frac{\nu_{(d)}^2 D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(d)}}\right) \boldsymbol{\mu}_{ij}^{(d)} \boldsymbol{\mu}_{ij}^{(d)^T} D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(d)}}\right)}{\boldsymbol{\mu}_{ij}^{(d)^T} D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(d)}}\right) \boldsymbol{\mu}_{ij}^{(d)}}$$

$$= D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(d)}}\right) - \frac{\nu_{(d)}^2}{1+\nu_{(d)}^2\mu_{ij.}^{(d)}}$$

Then, the j-th block of $\tilde{W}_i^{(d,d)}$ is given by

$$\begin{aligned}
\tilde{W}_{ij}^{(d,d)} &= \left(\tilde{V}_{ij}^{(d,d)} - \tilde{V}_{ij}^{(d,1-d)}\left(\tilde{V}_{ij}^{(1-d,1-d)}\right)^{-1}\tilde{V}_{ij}^{(1-d,d)}\right)^{-1} \\
&= \Bigg[\nu_{(d)}^2\boldsymbol{\mu}_{ij}^{(d)}\boldsymbol{\mu}_{ij}^{(d)^T} + D(\boldsymbol{\mu}_{ij}^{(d)}) - \theta^2\boldsymbol{\mu}_{ij}^{(d)}\boldsymbol{\mu}_{ij}^{(1-d)^T} \\
&\qquad \cdot\left(D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(1-d)}}\right) - \frac{\nu_{(1-d)}^2}{1+\nu_{(1-d)}^2\mu_{ij.}^{(1-d)}}\right)\boldsymbol{\mu}_{ij}^{(1-d)}\boldsymbol{\mu}_{ij}^{(d)^T}\Bigg]^{-1} \\
&= \left[D(\boldsymbol{\mu}_{ij}^{(d)}) + \left(\nu_{(d)}^2 - \theta^2\frac{\mu_{ij.}^{(1-d)}}{1+\nu_{(1-d)}^2\mu_{ij.}^{(1-d)}}\right)\boldsymbol{\mu}_{ij}^{(d)}\boldsymbol{\mu}_{ij}^{(d)^T}\right]^{-1} \\
&= D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(d)}}\right) - \frac{\nu_{(d)}^2 - \theta^2\frac{\mu_{ij.}^{(1-d)}}{1+\nu_{(1-d)}^2\mu_{ij.}^{(1-d)}}}{1+\left(\nu_{(d)}^2 - \theta^2\frac{\mu_{ij.}^{(1-d)}}{1+\nu_{(1-d)}^2\mu_{ij.}^{(1-d)}}\right)\mu_{ij.}^{(d)}}\mathbf{1}_{\boldsymbol{\mu}_{ij}^{(d)}}\mathbf{1}_{\boldsymbol{\mu}_{ij}^{(d)}}^T \\
&= D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(d)}}\right) - w_{ij}\left(\nu_{(d)}^2 + \nu_{(d)}^2\nu_{(1-d)}^2\mu_{ij.}^{(1-d)} - \theta\mu_{ij.}^{(1-d)}\right)\mathbf{1}_{\boldsymbol{\mu}_{ij}^{(d)}}\mathbf{1}_{\boldsymbol{\mu}_{ij}^{(d)}}^T \quad (3.37)
\end{aligned}$$

and the j-th block of $\tilde{W}_i^{(d,1-d)}$ is given by

$$\begin{aligned}
\tilde{W}_{ij}^{(d,1-d)} &= -\left(\tilde{V}_{ij}^{(d,d)}\right)^{-1}\tilde{V}_{ij}^{(d,1-d)}\tilde{W}_{ij}^{(1-d,1-d)} \\
&= -\left(D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(d)}}\right) - \frac{\nu_{(d)}^2}{1+\nu_{(d)}^2\mu_{ij.}^{(d)}}\right)\theta\boldsymbol{\mu}_{ij}^{(d)}\boldsymbol{\mu}_{ij}^{(1-d)} \\
&\qquad \cdot\left(D\left(\frac{1}{\boldsymbol{\mu}_{ij}^{(1-d)}}\right) - w_{ij}\left(1+\nu_{(1-d)}^2\nu_{(d)}^2\mu_{ij.}^{(d)} - \theta\mu_{ij.}^{(d)}\right)\right) \\
&= -\theta w_{ij}\mathbf{1}_{\boldsymbol{\mu}_{ij}^{(d)}}\mathbf{1}_{\boldsymbol{\mu}_{ij}^{(1-d)}}^T \quad (3.38)
\end{aligned}$$

Given this, it's possible to compute the values of $p_{ij}^{(d)}$ and $q_{ij}^{(d)}$:

$$\begin{aligned}
p_{ij}^{(d)} &= \boldsymbol{f}_{ij}^{(d)}\tilde{W}_i^{(d,d)}\boldsymbol{\mu}_i^{(d)} \\
&= \boldsymbol{\mu}_{ij}^{(d)}\tilde{W}_{ij}^{(d,d)}\boldsymbol{\mu}_{ij}^{(d)} \\
&= \mu_{ij.}^{(d)} - w_{ij}\left(\nu_{(d)}^2 + \nu_{(d)}^2\nu_{(1-d)}^2\mu_{ij.}^{(1-d)} - \theta\mu_{ij.}^{(1-d)}\right)\left(\mu_{ij.}^{(d)}\right)^2
\end{aligned}$$

$$= \mu_{ij.}^{(d)} w_{ij}(1 + \nu_{(1-d)}^2 \mu_{ij.}^{(1-d)}) \,,$$

which is the same as given in (3.18). The computations for $q_{ij}^{(d)}$, $P_{ij}^{(d)}$ and $Q_{ij}^{(d)}$ are analogous. Furthermore, it follows naturally that

$$p_i^{(d)} = \boldsymbol{\mu}_i^{(d)} \tilde{W}_i^{(d,d)} \boldsymbol{\mu}_i^{(d)} = \sum_{j=1}^{J_i} \boldsymbol{f}_{ij}^{(d)} \tilde{W}_i^{(d,d)} \boldsymbol{\mu}_i^{(d)} = \sum_{j=1}^{J_i} p_{ij}^{(d)} \,,$$

and similary for $P_i^{(d)}, q_i^{(d)}, Q_i^{(d)}$. This completes the derivation of the BLUP results given in Section 3.2.3.1.

3.3.2 Derivation of the bias-adjusted Pearson estimators

As noted in Section 3.2.3.2, the naïve dispersion parameters are biased due to variance shrinkage introduced by the BLUPs. This section uses general properties of the orthodox BLUP as well as the BLUP formulations computed in Section 3.3.1 to estimate the bias and show that, for fixed regression and baseline parameters, the corrections given in Section 3.2.3.2 are appropriate. The approach is similar to Ma (1999).

The computations make use of the following orthogonality properties of the BLUP:

$$\mathrm{Cov}\,(\hat{U} - U, \hat{U}) = 0 \,, \qquad \mathrm{Cov}\,(\hat{U} - U, \delta) = 0$$

that is, the difference between the frailty and its BLUP is orthogonal to both the predictor and the data.

3.3.2.1 Cluster-level dispersion parameters

The cluster-level dispersion parameter estimator is given by

$$\hat{\sigma}^2_{(d)} = \frac{1}{m}\sum_{i=1}^{m}\left\{(\hat{U}_i^{(d)} - 1)^2 + b_i^{(d)}\right\},$$

where

$$\begin{aligned} b_i^{(d)} &= \mathbb{E}\left[(\hat{U}_i^{(d)} - U_i^{(d)})^2\right] \\ &= \sigma^2_{(d)} w_i (1 + \sigma^2_{(1-d)} p_i^{(1-d)}) \end{aligned} \tag{3.39}$$

We will show that the bias correction $b_i^{(d)}$ leads to an unbiased estimator. Note that

$$\begin{aligned} \mathbb{E}\left[(U_i^{(d)} - 1)^2\right] &= \mathrm{Var}\,(U_i^{(d)}) \\ &= \mathrm{Var}\,(U_i^{(d)} - \hat{U}_i^{(d)} + \hat{U}_i^{(d)}) \\ &= \mathrm{Var}\,(U_i^{(d)} - \hat{U}_i^{(d)}) + \mathrm{Var}\,(\hat{U}_i^{(d)}) - 2\underbrace{\mathrm{Cov}\,(U_i^{(d)} - \hat{U}_i^{(d)}, \hat{U}_i^{(d)})}_{=0} \\ &= \mathbb{E}\left[(\hat{U}_i^{(d)} - U_i^{(d)})^2\right] + \mathrm{Var}\,(\hat{U}_i^{(d)}) \\ &= b_i^{(d)} + \mathbb{E}\left[(\hat{U}_i^{(d)} - 1)^2\right] . \end{aligned}$$

Reorganizing this expression shows that

$$\mathbb{E}\left[(\hat{U}_i^{(d)} - 1)^2\right] = \sigma^2_{(d)} + b_i^{(d)} , \tag{3.40}$$

so $b_i^{(d)}$ is the appropriate bias correction. To compute its value, we first use the form of the cluster-level BLUP given in (3.36) and the general result of (3.34), to compute

$$\mathrm{Var}\,(\hat{U}_i^{(d)}) = \sigma^4 w_i \left[(1 + \sigma^2_{(1-d)} p_i^{(1-d)}) p_i^{(d)} - \sigma^2 (1-d) {q_i^{(d)}}^2\right]$$

Reorganizing (3.40) and substituting this result allows $b_i^{(d)}$ to be written as

$$b_i^{(d)} = \mathrm{Var}\,(U_i^{(d)}) - \mathrm{Var}\,(\hat{U}_i^{(d)})$$

$$
\begin{aligned}
&= \sigma^2_{(d)} - \sigma^4 w_i \left[(1 + \sigma^2_{(1-d)} p_i^{(1-d)}) p_i^{(d)} - \sigma^2 (1-d) {q_i^{(d)}}^2 \right] \\
&= \sigma^2_{(d)} w_i \left(1 + \sigma^2_{(1-d)} p_i^{(1-d)} \right) ,
\end{aligned}
$$

which matches (3.39). It follows that the given estimator $\hat{\sigma}^2_{(d)}$ is unbiased for $\sigma^2_{(d)}$, assuming correct regression and baseline hazard parameter estimates. In implementation, this must be considered an update equation, since $b_i^{(d)}$ depends on the dispersion parameters as well. Thus, the previous iteration's dispersion parameter estimates may be used to compute the bias correction for the next iteration's estimators.

3.3.2.2 Subject-level dispersion parameters

The estimator for $\nu^2_{(d)}$ is derived analogously, though the computations are somewhat more involved. The proposed estimator is given by

$$
\hat{\nu}^2_{(d)} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{J_i} \sum_{j=1}^{J_i} \left\{ (\hat{U}_{ij}^{(d)} - \hat{U}_i^{(d)})^2 + b_{ij}^{(d)} \right\} ,
$$

where

$$
\begin{aligned}
b_{ij}^{(d)} &= \mathbb{E}\left[(\hat{U}_i^{(d)} - U_i^{(d)})^2 \right] + \mathbb{E}\left[(\hat{U}_{ij}^{(d)} - U_{ij}^{(d)})^2 \right] - 2 \left\{ \sigma^2_{(d)} - \mathrm{Cov}\left(\hat{U}_i^{(d)}, U_{ij}^{(d)} \right) \right\} \\
&= b_i^{(d)} + c_{ij}^{(d)} - 2(\sigma^2_{(d)} - z_{ij}^{(d)}) , \qquad (3.41)
\end{aligned}
$$

where we denote $c_{ij}^{(d)} = \mathbb{E}\left[(\hat{U}_{ij}^{(d)} - U_{ij}^{(d)})^2 \right]$ and $z_{ij}^{(d)} = \mathrm{Cov}\left(\hat{U}_i^{(d)}, U_{ij}^{(d)} \right)$.

To justify this correction, compute the expectation of the Pearson-type estimator directly:

$$
\mathbb{E}\left[(\hat{U}_{ij}^{(d)} - \hat{U}_i^{(d)})^2 \right] = \underbrace{\mathrm{Var}\left(\hat{U}_{ij}^{(d)} \right)}_{[1]} + \underbrace{\mathrm{Var}\left(\hat{U}_i^{(d)} \right)}_{[2]} - 2 \underbrace{\mathrm{Cov}(\hat{U}_{ij}^{(d)}, \hat{U}_i^{(d)})}_{[3]} . \qquad (3.42)
$$

To compute component [1], note that

$$
\mathrm{Var}\left(U_{ij}^{(d)} \right) = \mathrm{Var}\left(U_{ij}^{(d)} - \hat{U}_{ij}^{(d)} + \hat{U}_{ij}^{(d)} \right)
$$

$$
\begin{aligned}
&= \operatorname{Var}(U_{ij}^{(d)} - \hat{U}_{ij}^{(d)}) + \operatorname{Var}(\hat{U}_{ij}^{(d)}) + 2\underbrace{\operatorname{Cov}(U_{ij}^{(d)} - \hat{U}_{ij}^{(d)}, \hat{U}_{ij}^{(d)})}_{=0} \\
&= c_{ij}^{(d)} + \operatorname{Var}(\hat{U}_{ij}^{(d)}) ,
\end{aligned}
$$

and therefore

$$
\operatorname{Var}(\hat{U}_{ij}^{(d)}) = \operatorname{Var}(U_{ij}^{(d)}) - c_{ij}^{(d)} = \sigma_{(d)}^2 + \nu^2(d) - c_{ij}^{(d)} . \tag{3.43}
$$

Component [2] follows from equation (3.40):

$$
\operatorname{Var}(\hat{U}_{i}^{(d)}) = \sigma_{(d)}^2 - b_{i}^{(d)} .
$$

Component [3] can be simplified as

$$
\operatorname{Cov}(\hat{U}_{ij}^{(d)}, \hat{U}_{i}^{(d)}) = \underbrace{\operatorname{Cov}(\hat{U}_{ij}^{(d)} - U_{ij}^{(d)}, \hat{U}_{i})}_{=0} + \operatorname{Cov}(\hat{U}_{i}^{(d)}, U_{ij}^{(d)}) ,
$$

and recall that we denoted $\operatorname{Cov}(\hat{U}_{ij}^{(d)}, \hat{U}_{i}^{(d)}) = z_{ij}^{(d)}$. Therefore, substituting these results back into (3.42) gives

$$
\begin{aligned}
\mathbb{E}\left[(\hat{U}_{ij}^{(d)} - \hat{U}_{i}^{(d)})^2\right] &= \operatorname{Var}(\hat{U}_{ij}^{(d)}) + \operatorname{Var}(\hat{U}_{i}^{(d)}) - 2\operatorname{Cov}(\hat{U}_{ij}^{(d)}, \hat{U}_{i}^{(d)}) \\
&= (\sigma_{(d)}^2 + \nu_{(d)}^2 - c_{ij}^{(d)}) + (\sigma_{(d)}^2 - b_{i}^{(d)}) - 2z_{ij}^{(d)} \\
&= \nu_{(d)}^2 - b_{ij}^{(d)} .
\end{aligned}
$$

where $b_{ij}^{(d)}$ matches eq. (3.41).

It remains to compute the values of $c_{ij}^{(d)}$ and $z_{ij}^{(d)}$. In order to compute the value of $c_{ij}^{(d)}$, rearrange eq. (3.40), to yield

$$
\begin{aligned}
c_{ij}^{(d)} &= \operatorname{Var}(U_{ij}^{(d)}) - \operatorname{Var}(\hat{U}_{ij}^{(d)}) \\
&= \sigma_{(d)}^2 + \nu_{(d)}^2 - \Big[\underbrace{\operatorname{Cov}(\hat{U}_{ij}^{(d)} - U_{ij}^{(d)}, \hat{U}_{ij}^{(d)})}_{=0} + \operatorname{Cov}(\hat{U}_{ij}^{(d)}, U_{ij}^{(d)})\Big] \\
&= \sigma_{(d)}^2 + \nu_{(d)}^2 - \operatorname{Cov}(\hat{U}_{ij}^{(d)}, U_{ij}^{(d)}) .
\end{aligned} \tag{3.44}
$$

Substituting the formula for $\hat{U}_{ij}^{(d)}$ in (3.16), the covariance term in this equation may be computed as

$$\begin{aligned}\text{Cov}\left(\hat{U}_{ij}^{(d)}, U_{ij}^{(d)}\right) &= w_{ij}\left(1+\nu_{(1-d)}^2\mu_{ij.}^{(1-d)}\right)\text{Cov}\left(\hat{U}_{i}^{(d)}, U_{ij}^{(d)}\right) \\ &\quad -\left(\nu_{(d)}^2 q_{ij}^{(1-d)}+\theta p_{ij}^{(1-d)}\right)\text{Cov}\left(\hat{U}_{i}^{(1-d)}, U_{ij}^{(d)}\right) \\ &\quad +\nu_{(d)}^2\text{Cov}\left(P_{ij}^{(d)}+Q_{ij}^{(1-d)}, U_{ij}^{(d)}\right)+\theta\text{Cov}\left(P_{ij}^{(1-d)}+Q_{ij}^{(d)}, U_{ij}^{(d)}\right)\end{aligned} \tag{3.45}$$

Further simplification of this expression relies on the following results:

$$\begin{aligned}&\text{Cov}\left(P_{it}^{(d)}, U_{ij}^{(d)}\right) = (\sigma_{(d)}^2+\mathbf{1}_{(t,j)}\nu_{(d)}^2)p_{it}^{(d)} \qquad &&\text{Cov}\left(P_{it}^{(d)}, U_{ij}^{(1-d)}\right) = \mathbf{1}_{(t,j)}\theta p_{it}^{(d)} \\ &\text{Cov}\left(Q_{it}^{(d)}, U_{ij}^{(d)}\right) = (\sigma_{(d)}^2+\mathbf{1}_{(t,j)}\nu_{(d)}^2)q_{it}^{(d)} \qquad &&\text{Cov}\left(Q_{it}^{(d)}, U_{ij}^{(1-d)}\right) = \mathbf{1}_{(t,j)}\theta q_{it}^{(d)}\end{aligned} \tag{3.46}$$

This allows computation of required covariance terms as

$$\begin{aligned}z_{ij}^{(d)} &= \text{Cov}\left(\hat{U}_{i}^{(d)}, U_{ij}^{(d)}\right) \\ &= \sigma_{(d)}^2 w_i\big[(1+\sigma_{(1-d)}^2 p_i^{(1-d)})\text{Cov}\left(P_i^{(d)}+Q_i^{(1-d)}, U_{ij}^{(d)}\right) \\ &\qquad -\sigma_{(1-d)}^2 q_i^{(d)}\text{Cov}\left(P_i^{(1-d)}+Q_i^{(d)}, U_{ij}^{(d)}\right)\big]\,, \\ y_{ij}^{(d)} &= \text{Cov}\left(\hat{U}_{i}^{(d)}, U_{ij}^{(1-d)}\right) \\ &= \sigma_{(d)}^2 w_i\big[(1+\sigma_{(1-d)}^2 p_i^{(1-d)})\text{Cov}\left(P_i^{(d)}+Q_i^{(1-d)}, U_{ij}^{(1-d)}\right) \\ &\qquad -\sigma_{(1-d)}^2 q_i^{(d)}\text{Cov}\left(P_i^{(1-d)}+Q_i^{(d)}, U_{ij}^{(1-d)}\right)\big]\,,\end{aligned}$$

and substituting the results of (3.46) gives

$$\begin{aligned}z_{ij}^{(d)} &= \sigma_{(d)}^2 w_i\big[(1+\sigma_{(1-d)}^2 p_i^{(1-d)})(\sigma_{(d)}^2 p_i^{(d)}+\nu_{(d)}^2 p_{ij}^{(d)}+\theta q_{ij}^{(d)}) \\ &\qquad -\sigma_{(1-d)}^2 q_i^{(d)}(\sigma_{(d)}^2 q_i^{(d)}+\nu_{(d)}^2 q_{ij}^{(d)}+\theta p_{ij}^{(1-d)})\big]\,, \\ y_{ij}^{(d)} &= \sigma_{(d)}^2 w_i\big[(1+\sigma_{(1-d)}^2 p_i^{(1-d)})(\theta p_{ij}^{(d)}+\nu_{(1-d)}^2 q_{ij}^{(d)}) \\ &\qquad -\sigma_{(1-d)}^2 q_i^{(d)}(\theta q_{ij}^{(d)}+\nu_{(1-d)}^2 p_{ij}^{(1-d)}-1)\big]\,.\end{aligned}$$

Substituting this back into (3.45) gives the desired covariance expression as

$$\begin{aligned}\text{Cov}\left(\hat{U}_{ij}^{(d)}, U_{ij}^{(d)}\right) = w_{ij}\left(1+\nu_{(1-d)}^2\mu_{ij.}^{(1-d)}\right)z_{ij}^{(d)} - \left(\nu_{(d)}^2 q_{ij}^{(1-d)} + \theta p_{ij}^{(1-d)}\right)y_{ij}^{(1-d)} \\ -\left(\sigma_{(d)}^2+\nu_{(d)}^2\right)\left(\nu_{(d)}^2 p_{ij}^{(d)} - \theta q_{ij}^{(d)}\right) + \theta\left(\nu_{(d)}^2 q_{ij}^{(1-d)} + \theta p_{ij}^{(1-d)}\right) .\end{aligned} \tag{3.47}$$

Substituting this into (3.44) and simplifying further gives

$$c_{ij}^{(d)} = (\nu_{(d)}^2 p_{ij}^{(d)} + \theta q_{ij}^{(1-d)} - 1)(z_{ij}^{(d)} - \sigma_{(d)}^2 - \nu_{(d)}^2) + (\nu_{(d)}^2 q_{ij}^{(1-d)} + \theta p_{ij}^{(1-d)})(y_{ij}^{(1-d)} - \theta) .$$

The bias correction again depends on the parameters themselves. At each iteration, estimates from the previous iteration can be used to update the parameter estimates.

3.3.2.3 Frailty covariance parameter

The proposed estimator for the frailty covariance is:

$$\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{J_i}\sum_{j=1}^{J_i}\left\{(\hat{U}_{ij}^{(0)} - \hat{U}_i^{(0)})(\hat{U}_{ij}^{(1)} - \hat{U}_i^{(1)}) + b_{ij}^{(*)}\right\},$$

where

$$\begin{aligned} b_{ij}^{(*)} &= \mathbb{E}\left[(\hat{U}_{ij}^{(0)} - U_{ij}^{(0)})(\hat{U}_{ij}^{(1)} - U_{ij}^{(1)})\right] \\ &\quad +\text{Cov}\left(\hat{U}_{ij}^{(0)}, U_{ij}^{(1)}\right) + \text{Cov}\left(\hat{U}_{ij}^{(1)}, U_{ij}^{(0)}\right) - \text{Cov}\left(\hat{U}_i^{(0)}, \hat{U}_i^{(1)}\right) \\ &= c'_{ij} + y_{ij}^{(0)} + y_{ij}^{(1)} - \sigma_{(0)}^2\sigma_{(1)}^2 w_i q_i^{(0)} , \end{aligned}$$

where we denote $c'_{ij} = \mathbb{E}\left[(\hat{U}_{ij}^{(0)} - U_{ij}^{(0)})(\hat{U}_{ij}^{(1)} - U_{ij}^{(1)})\right]$. To justify this, note that the expectation of the naïve Pearson-type estimator can be computed as:

$$\mathbb{E}\left[(\hat{U}_{ij}^{(0)} - \hat{U}_i^{(0)})(\hat{U}_{ij}^{(1)} - \hat{U}_i^{(1)})\right] = \underbrace{\text{Cov}\left(\hat{U}_{ij}^{(0)}, \hat{U}_{ij}^{(1)}\right)}_{[1]} - \underbrace{\text{Cov}\left(\hat{U}_{ij}^{(0)}, \hat{U}_i^{(1)}\right)}_{[2]}$$

$$- \underbrace{\mathrm{Cov}\,(\hat{U}_{ij}^{(1)}, \hat{U}_{i}^{(0)})}_{[3]} + \underbrace{\mathrm{Cov}\,(\hat{U}_{i}^{(0)}, \hat{U}_{i}^{(1)})}_{[4]} \quad (3.48)$$

To compute component [1], note that

$$\begin{aligned}
\mathrm{Cov}\,(U_{ij}^{(d)}, U_{ij}^{(1-d)}) &= \mathrm{Cov}\,(U_{ij}^{(d)} - \hat{U}_{ij}^{(d)} + \hat{U}_{ij}^{(d)}, U_{ij}^{(1-d)}) \\
&= \mathrm{Cov}\,(U_{ij}^{(d)} - \hat{U}_{ij}^{(d)}, U_{ij}^{(1-d)} + \mathrm{Cov}\,(\hat{U}_{ij}^{(d)}, U_{ij}^{(1-d)}) \\
&= \mathrm{Cov}\,(\hat{U}_{ij}^{(d)} - U_{ij}^{(d)}, \hat{U}_{ij}^{(1-d)} - U_{ij}^{(1-d)}) \\
&\quad + \underbrace{\mathrm{Cov}\,(U_{ij}^{(d)} - \hat{U}_{ij}^{(d)}, \hat{U}_{ij}^{(1-d)})}_{=0} \\
&\quad + \mathrm{Cov}\,(\hat{U}_{ij}^{(d)}, \hat{U}_{ij}^{(1-d)}) \\
&\quad + \underbrace{\mathrm{Cov}\,(\hat{U}_{ij}^{(d)}, U_{ij}^{(1-d)} - \hat{U}_{ij}^{(1-d)})}_{=0} \\
&= c'_{ij} + \mathrm{Cov}\,(\hat{U}_{ij}^{(d)}, \hat{U}_{ij}^{(1-d)})\,,
\end{aligned}$$

and therefore, component [1] is given by $\theta - c'_{ij}$.

The second and third components are given by $y_{ij}^{(0)}$ and $y_{ij}^{(1)}$ respectively. For component [4], one can compute

$$\begin{aligned}
\mathrm{Cov}\,(\hat{U}_{i}^{(1)}, \hat{U}_{i}^{(0)}) &= \mathrm{Cov}\,(\hat{U}_{i}^{(1)}, U_{i}^{(0)}) \\
&= \sigma_{(1)}^2 w_i \big[(1 + \sigma_{(0)}^2 p_i^{(0)})\mathrm{Cov}\,(Q_i^{(1)} + P_i^{(1)}, U_i^{(0)}) \\
&\qquad -\sigma_{(0)}^2 q_i^{(0)} \mathrm{Cov}\,(P_i^{(0)} + Q_i^{(0)}, U_i^{(0)})\big] \\
&= \sigma_{(1)}^2 w_i \big[(1 + \sigma_{(0)}^2 p_i^{(0)})\sigma_{(0)}^2 q_i^{(0)} - \sigma_{(0)}^2 q_i^{(0)} (\sigma_{(0)}^2 p_i^{(0)})\big] \\
&= \sigma_{(0)}^2 \sigma_{(1)}^2 w_i q_i^{(0)}
\end{aligned}$$

Substituting these results into eq. (3.48) shows that

$$\mathbb{E}\left[(\hat{U}_{ij}^{(0)} - \hat{U}_{i}^{(0)})(\hat{U}_{ij}^{(1)} - \hat{U}_{i}^{(1)})\right] = \theta - b_{ij}^{(*)}\,,$$

and therefore $b_{ij}^{(*)}$ is the required bias correction.

It remains to compute the value of c'_{ij}. Rearranging eq. (3.49) gives

$$c'_{ij} = \theta - \mathrm{Cov}\,(\hat{U}_{ij}^{(d)}, \hat{U}_{ij}^{(1-d)})\ .$$

The covariance term in this expression can be computed by arbitrarily breaking the symmetry, as

$$\begin{aligned} c'_{ij} &= \theta - \mathrm{Cov}\,(\hat{U}_{ij}^{(0)}, U_{ij}^{(1)}) \\ &= \theta - (1 - \nu_{(0)}^2 p_{ij}^{(0)} - \theta q_{ij}^{(0)}) y_{ij}^{(0)} - (\nu_{(0)}^2 q_{ij}^{(0)} + \theta p_{ij}^{(1)}) z_{ij}^{(1)} \\ &\quad + (\sigma_{(1)}^2 + \nu_{(1)}^2)(\nu_{(0)}^2 q_{ij}^{(0)} + \theta p_{ij}^{(1)}) + \theta(\nu_{(0)}^2 p_{ij}^{(0)} + \theta q_{ij}^{(0)})\ , \end{aligned}$$

where we used the result of eq. (3.47).

We have thus shown that all estimators proposed in the summary are unbiased for their respective parameters, given accurate regression parameter and frailty estimates.

3.3.3 Construction of the Godambe matrix

This section details the construction of the Godambe matrix, whose form was given in Section 3.2.4 for purposes of obtaining standard errors. We follow to a large extent the methods presented in Ma (1999).

3.3.3.1 Additional notation

As before, denote the vector of regression parameters and baseline parameters for process (d) by $\gamma^{(d)} = [\log \alpha^{(d)}, \beta^{(d)}]$, and correspondingly, denote

$$x_{rijks}^{(d)} = (0, \ldots, 0, 1, 0, \ldots, Z_{ij}^T)^T\ ,$$

for $Y^{(d)}_{rijks} = 1$, where the 1 is in the position corresponding to $\alpha^{(d)}_{rs}$. In this notation,

$$\mu^{(d)}_{rijks} = e^{\gamma^{(d)} x^{(d)}_{rijks}} A^{(d)}_{rs}(T^{(d)}_{ijk}) \ ,$$

subject to the accuracy of the auxiliary Poisson model. This allows the conditional likelihood of (3.10) to be written as

$$\ell(\gamma|U^{(*)}_{*}) = \sum_{d,r,i,j,k,s} Y^{(d)}_{rijks} \Bigg[\delta^{(d)}_{rijks} \left(\log U^{(d)}_{ij} + \gamma^{(d)} x^{(d)}_{rijks} \right) - U^{(d)}_{ij} e^{\gamma^{(d)} x^{(d)}_{rijks}} A^{(d)}_{rs}(T^{(d)}_{ijk}) \Bigg] \ , \quad (3.49)$$

In addition, for purposes of the proofs in Section 3.3.3.2, denote vectors of the products $U^{(d)}_{ij}\mu^{(d)}_{rijks}$ as

$$\begin{aligned} U\mu^{(d)}_{i} &= \left(U^{(d)}_{i1}\mu^{(d)}_{1i111}, \ldots U^{(d)}_{i1}\mu^{(d)}_{pi1M_{i1}K_p}, \ldots, U^{(d)}_{iJ_i}\mu^{(d)}_{1iJ_i11}, \ldots U^{(d)}_{iJ_i}\mu^{(d)}_{piJ_iM_{i1}K_p} \right) \\ \hat{U}\mu^{(d)}_{i} &= \left(\hat{U}^{(d)}_{i1}\mu^{(d)}_{1i111}, \ldots \hat{U}^{(d)}_{i1}\mu^{(d)}_{pi1M_{i1}K_p}, \ldots, \hat{U}^{(d)}_{iJ_i}\mu^{(d)}_{1iJ_i11}, \ldots \hat{U}^{(d)}_{iJ_i}\mu^{(d)}_{piJ_iM_{i1}K_p} \right) \ , \end{aligned}$$

and vectors of related variables by boldfaced symbols,

$$\begin{aligned} \boldsymbol{\mu}_i &= \begin{bmatrix} \boldsymbol{\mu}^{(0)}_i \\ \boldsymbol{\mu}^{(1)}_i \end{bmatrix} \qquad \boldsymbol{U\mu_i} = \begin{bmatrix} U\mu^{(0)}_i \\ U\mu^{(1)}_i \end{bmatrix} \\ \boldsymbol{\delta}_i &= \begin{bmatrix} \boldsymbol{\delta}^{(0)}_i \\ \boldsymbol{\delta}^{(1)}_i \end{bmatrix} \qquad \boldsymbol{\hat{U}\mu_i} = \begin{bmatrix} \hat{U}\mu^{(0)}_i \\ \hat{U}\mu^{(1)}_i \end{bmatrix} \ , \end{aligned} \quad (3.50)$$

and the data matrix for each cluster i as

$$\boldsymbol{X}_i = \begin{bmatrix} X^{(0)}_i & 0 \\ 0 & X^{(1)}_i \end{bmatrix} \ . \quad (3.51)$$

3.3.3.2 Structure of the Godambe matrix

This notation allows the Godambe matrix to more easily be expressed in terms of the sensitivity and variability matrices. In this section, we will show that the Godambe matrix is in fact simply the negative of the sensitivity matrix.

The gradient of the likelihood in (3.49) represents a set of estimating equations

$$\psi(\gamma^{(0)}, \gamma^{(1)}) = \sum_{i=1}^{m} \psi_i(\gamma^{(0)}, \gamma^{(1)}) = 0$$

where

$$\psi_i(\gamma^{(0)}, \gamma^{(1)}) = \begin{bmatrix} \sum_{r,j,k,s} Y^{(0)}_{rijks} x^{(0)}_{rijks} \left(\delta^{(0)}_{rijks} - \hat{U}^{(0)}_{ij} e^{\gamma^{(0)} x^{(0)}_{rijks}} A^{(0)}_{rs}(T^{(0)}_{ijk}) \right) \\ \sum_{r,j,k,s} Y^{(1)}_{rijks} x^{(1)}_{rijks} \left(\delta^{(1)}_{rijks} - \hat{U}^{(1)}_{ij} e^{\gamma^{(1)} x^{(1)}_{rijks}} A^{(1)}_{rs}(T^{(1)}_{ijk}) \right) \end{bmatrix} .$$

Using the notation of (3.50) and (3.51), it can also be written as

$$\psi_i(\gamma^{(0)}, \gamma^{(1)}) = \boldsymbol{X}_i^T (\boldsymbol{\delta}_i - \hat{\boldsymbol{U}} \boldsymbol{\mu}_i) . \tag{3.52}$$

The Godambe matrix can be expressed in terms of the sensitivity and variability matrices

$$G(\gamma^{(0)}, \gamma^{(1)}) = S(\gamma^{(0)}, \gamma^{(1)}) V(\gamma^{(0)}, \gamma^{(1)})^{-1} S(\gamma^{(0)}, \gamma^{(1)}) ,$$

where the sensitivity and variability matrices are defined as

$$\begin{aligned} S(\gamma^{(0)}, \gamma^{(1)}) &= \sum_{i=1}^{m} S_i(\gamma^{(0)}, \gamma^{(1)}) = \sum_{i=1}^{m} \mathbb{E} \left[\frac{\partial \psi_i(\gamma^{(0)}, \gamma^{(1)})}{\partial [{\gamma^{(0)}}^T, {\gamma^{(1)}}^T]} \right] , \\ V(\gamma^{(0)}, \gamma^{(1)}) &= \sum_{i=1}^{m} V_i(\gamma^{(0)}, \gamma^{(1)}) = \sum_{i=1}^{m} \mathbb{E} \left[\psi_i(\gamma^{(0)}, \gamma^{(1)}) \psi_i^T(\gamma^{(0)}, \gamma^{(1)}) \right] . \end{aligned}$$

In analogy to Ma (1999), we will show in the following that the estimating equation, sensitivity and variability can be written as

$$\psi_i(\gamma^{(0)}, \gamma^{(1)}) = \boldsymbol{X}_i^T D(\mathbb{E}[\boldsymbol{\delta}_i]) \mathrm{Var}(\boldsymbol{\delta}_i)^{-1} (\boldsymbol{\delta}_i - \boldsymbol{\mu}_i) \tag{3.53}$$

$$S_i(\gamma^{(0)}, \gamma^{(1)}) = -\boldsymbol{X}_i^T D(\mathbb{E}[\boldsymbol{\delta}_i]) \mathrm{Var}(\boldsymbol{\delta}_i)^{-1} D(\mathbb{E}[\boldsymbol{\delta}_i]) \boldsymbol{X}_i \tag{3.54}$$

$$V_i(\gamma^{(0)}, \gamma^{(1)}) = -S_i(\gamma^{(0)}, \gamma^{(1)}) . \tag{3.55}$$

In order to prove (3.53), we can rewrite (3.52) by noting that by (3.14)

$$\hat{\boldsymbol{U}} \boldsymbol{\mu}_i = \mathbb{E}[\boldsymbol{\delta}_i] + \mathrm{Cov}(\boldsymbol{U} \boldsymbol{\mu}_i, \boldsymbol{\delta}_i) \mathrm{Var}(\boldsymbol{\delta}_i)^{-1} (\boldsymbol{\delta}_i - \mathbb{E}[\boldsymbol{\delta}_i]) ,$$

and also,

$$\begin{aligned}
\mathrm{Var}\left(\boldsymbol{\delta}_i^{(d)}\right) &= \mathrm{Cov}\left(U\mu_i^{(d)}, \boldsymbol{\delta}_i^{(d)}\right) + D(\boldsymbol{\mu}_i^{(d)}) \\
\mathrm{Cov}\left(\boldsymbol{\delta}_i^{(0)}, \boldsymbol{\delta}_i^{(1)}\right) &= \mathrm{Cov}\left(U\mu_i^{(0)}, \boldsymbol{\delta}_i^{(1)}\right)
\end{aligned}$$

so the overall covariance matrix can be written as

$$\mathrm{Var}\left(\boldsymbol{\delta}_i\right) = \mathrm{Cov}\left(\boldsymbol{U}\boldsymbol{\mu}_i, \boldsymbol{\delta}_i\right) + D(\boldsymbol{\mu}_i) \ .$$

Therefore, the last term of the estimating equation matrix form of (3.52) can be written as

$$\begin{aligned}
(\boldsymbol{\delta}_i - \hat{U}\mu_i) &= \boldsymbol{\delta}_i - \mathbb{E}\left[\boldsymbol{\delta}_i\right] - \mathrm{Cov}\left(\boldsymbol{U}\boldsymbol{\mu}_i, \boldsymbol{\delta}_i\right)\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}(\boldsymbol{\delta}_i - \mathbb{E}\left[\boldsymbol{\delta}_i\right]) \\
&= \left(\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1} - \mathrm{Cov}\left(\boldsymbol{U}\boldsymbol{\mu}_i, \boldsymbol{\delta}_i\right)\right)\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}(\boldsymbol{\delta}_i - \mathbb{E}\left[\boldsymbol{\delta}_i\right]) \\
&= D(\boldsymbol{\mu}_i)\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}(\boldsymbol{\delta}_i - \mathbb{E}\left[\boldsymbol{\delta}_i\right]) \ .
\end{aligned}$$

Substituting this into (3.52) yields (3.53).

The proofs of (3.54) and (3.55) follow immediately from the results of Ma (1999), but the arguments are repeated here for completeness. Given the form of $\psi_i(\gamma^{(0)}, \gamma^{(1)})$ from (3.53), the variability can be computed explicitly as

$$\begin{aligned}
V_i(\gamma^{(0)}, \gamma^{(1)}) &= \mathrm{Var}\left(\psi_i(\gamma^{(0)}, \gamma^{(1)})\right) \\
&= \mathrm{Var}\left(\boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}(\boldsymbol{\delta}_i - \mathbb{E}\left[\boldsymbol{\delta}_i\right])\right) \\
&= \boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}\mathrm{Var}\left(\boldsymbol{\delta}_i\right)\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}D(\boldsymbol{\mu}_i)\boldsymbol{X}_i \\
&= \boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}D(\boldsymbol{\mu}_i)\boldsymbol{X}_i \ ,
\end{aligned}$$

proving (3.54). The sensitivity matrix can be computed explicitly as well,

$$\begin{aligned}
S_i(\gamma^{(0)}, \gamma^{(1)}) &= \mathbb{E}\left[\frac{\partial\psi_i(\gamma^{(0)}, \gamma^{(1)})}{\partial[\gamma^{(0)^T}, \gamma^{(1)^T}]}\right] \\
&= \mathbb{E}\left[\frac{\partial}{\partial[\gamma^{(0)^T}, \gamma^{(1)^T}]}\boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\left(\boldsymbol{\delta}_i\right)^{-1}(\boldsymbol{\delta}_i - \mathbb{E}\left[\boldsymbol{\delta}_i\right])\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\left(\frac{\partial}{\partial[\gamma^{(0)^T},\gamma^{(1)^T}]}\boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\,(\boldsymbol{\delta}_i)^{-1}\right)(\boldsymbol{\delta}_i - \mathbb{E}\,[\boldsymbol{\delta}_i])\right] \\
&\quad +\mathbb{E}\left[\left(\boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\,(\boldsymbol{\delta}_i)^{-1}\right)\left(\frac{\partial}{\partial[\gamma^{(0)^T},\gamma^{(1)^T}]}(\boldsymbol{\delta}_i - \mathbb{E}\,[\boldsymbol{\delta}_i])\right)\right] \\
&= \left(\frac{\partial}{\partial[\gamma^{(0)^T},\gamma^{(1)^T}]}\boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\,(\boldsymbol{\delta}_i)^{-1}\right)\mathbb{E}\,[(\boldsymbol{\delta}_i - \mathbb{E}\,[\boldsymbol{\delta}_i])] \\
&\quad +\left(\boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\,(\boldsymbol{\delta}_i)^{-1}\right)\left(\frac{\partial}{\partial[\gamma^{(0)^T},\gamma^{(1)^T}]}\boldsymbol{\mu}_i\right) \\
&= 0 - \boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\mathrm{Var}\,(\boldsymbol{\delta}_i)^{-1} D(\boldsymbol{\mu}_i)\boldsymbol{X}_i \\
&= -V_i(\gamma^{(0)},\gamma^{(1)})\ ,
\end{aligned}
$$

proving (3.55).

This implies that the Godambe matrix of (3.3.3.2) is simply given by the sensitivity matrix, so that

$$
G(\gamma^{(0)},\gamma^{(1)}) = -S(\gamma^{(0)},\gamma^{(1)}) = -\sum_{i=1}^{m} S_i(\gamma^{(0)},\gamma^{(1)})\ .
$$

3.3.3.3 Computation of standard errors

The sensitivity matrix is computed by substituting (3.33) for the inverse variance matrix in (3.54). That is,

$$
\begin{aligned}
S_i(\gamma^{(0)},\gamma^{(1)}) &= -\boldsymbol{X}_i^T D(\boldsymbol{\mu}_i) W_i^{-1} D(\boldsymbol{\mu}_i)\boldsymbol{X}_i \\
&= -S_{i1} + w_i\left\{\sigma_{(0)}^2(1+\sigma_{(1)}^2 p_i^{(1)}) S_{i2}^{(0)^{\otimes 2}}\right. \\
&\qquad \left. -\sigma_{(0)}^2\sigma_{(1)}^2 q_i^{(0)} S_{i2}^{(0)}\nabla S_{i2}^{(1)} + \sigma_{(1)}^2(1+\sigma_{(0)} p_i^{(0)}) S_{i2}^{(1)^{\otimes 2}}\right\}\ ,
\end{aligned}
$$

where

$$
\begin{aligned}
S_{i1} &= \boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\tilde{W}_i^{-1} D(\boldsymbol{\mu}_i)\boldsymbol{X}_i \\
S_{i2}^{(d)} &= \boldsymbol{X}_i^T D(\boldsymbol{\mu}_i)\begin{bmatrix}\tilde{W}_i^{(0,d)}\boldsymbol{\mu}_i^{(d)} \\ \tilde{W}_i^{(1,d)}\boldsymbol{\mu}_i^{(d)}\end{bmatrix}
\end{aligned}
$$

Component S_{i1} can be computed by separating $\tilde{W}_i$ into quadrants, so that

$$S_{i1} = \begin{bmatrix} S_{i1}^{(0,0)} & S_{i1}^{(0,1)} \\ S_{i1}^{(1,0)} & S_{i1}^{(1,1)} \end{bmatrix}, \quad \text{where} \quad S_{i1}^{(d,e)} = \boldsymbol{X}_i^{(d)^T} D(\boldsymbol{\mu}_i^{(d)}) \tilde{W}_i^{(d,e)} D(\boldsymbol{\mu}_i^{(e)}) \boldsymbol{X}_i^{(e)} .$$

Moreover, each of the quadrants of $\tilde{W}_i$ has a block form given by eq (3.37) and (3.38) for diagonal and off-diagonal quadrants respectively. Substituting these expressions leads directly to the results in Section 3.2.4.

Standard errors for $\gamma^{(0)}, \gamma^{(1)}$ can be found from the sensitivity matrix as

$$\sigma_{\gamma^{(0)},\gamma^{(1)}} = \text{diag}\left[\left(-S(\gamma^{(0)}, \gamma^{(1)})\right)^{1/2}\right] .$$

3.4 Computational considerations

In this section, we discuss the implementation of the method and present the relevant computational considerations. Section 3.4.1 discusses the problem of finding adequate initial values for the algorithm, Section 3.4.2 suggests a method for choosing the number of discretization intervals $K^{(d)}$, and Section 3.4.3 shows how to select the interval boundary points $a_{rs}^{(d)}$. Section 3.4.4 gives examples of the computer time required to fit models with different sample sizes and discretizations in order to present the relative computational effort.

3.4.1 Obtaining initial values

This section discusses the problem of obtaining initial values for all the parameters that need to be estimated by the algorithm. Initial values need to be found for the regression parameter estimates $\hat{\beta}^{(d)}$, the baseline hazard parameters $\hat{\alpha}_{rs}^{(d)}$ for

$r = 1 \ldots p$ and $s = 1 \ldots K_r^{(d)}$, the set of cluster- and subject-level frailties $\hat{U}_i^{(d)}, \hat{U}_{ij}^{(d)}$, and the dispersion parameters $\hat{\sigma}^2_{(d)}, \hat{\nu}^2_{(d)}, \hat{\theta}$.

Many statistical software packages offer facilities for fitting random effects survival models, such as the `coxph` function in `R` (see Therneau and Grambsch, 2000). Fitting four distinct frailty survival models to the data allows reasonable initial values to be obtained: Fitting a model to only the recurrent event data for event d with cluster-level frailties gives estimates for the cluster-level frailties $\hat{U}_i^{(d)}$, as well as the frailty variance $\hat{\sigma}^2_{(d)}$. Fitting a model to recurrent event data for event type d with subject-level frailties gives estimates for $\hat{U}_{ij}^{(d)}$ and the marginal frailty variance $\hat{\sigma}^2_{(d)} + \hat{\nu}^2_{(d)}$. Thus an estimate for $\hat{\nu}^2_{(d)}$ can be obtained by subtracting the previous estimate for $\hat{\sigma}^2_{(d)}$ from the frailty variance returned by the fitted model. An estimate for $\hat{\theta}$ is given by the empirical covariance between the estimated values of $\hat{U}_{ij}^{(0)}$ and $\hat{U}_{ij}^{(1)}$.

Initial values for the regression parameters $\hat{\beta}^{(d)}$ can be obtained from the fitted survival models with subject-level frailties. The baseline hazard parameters can subsequently be estimated via (3.20).

Note that it is not guaranteed that these initial values satisfy the covariance structure in (3.1) through (3.4). In particular, the estimated $\hat{\nu}^2_{(0)}, \hat{\nu}^2_{(1)}, \hat{\theta}$ may not form a proper covariance matrix – in these cases, the estimate of $\hat{\theta}$ may need to be trimmed. Furthermore, it is necessary for computational purposes that the initial values for the dispersion parameters be nonzero, and it is therefore necessary to place a lower cap on the initial dispersion parameter estimates.

3.4.2 Choosing the number of discretization intervals

Prior to estimation, the number of breakpoints $K_r^{(d)}$ and their positions $a_{rs}^{(d)}$ in the discrete baseline hazard of (3.6) must be chosen for each d, r, s. The choice of $K_r^{(d)}$ represents a tradeoff between precision and numerical stability and speed – more accurate estimation of the baseline hazard may improve estimates of the regression coefficients and other parameters, but the addition of these nuisance parameters increases the computational complexity and may lead to instability in the estimation algorithm. The simulation results of Section 3.6.2 suggest that for very large samples, a relatively coarse level of discretization can yield good results with relatively little computational effort. Setting the number of breakpoints equal to the number of observed events in each stratum amounts to not discretizing the baseline hazard, similar to the methods employed by Ma et al. (2001).

Denote by $\phi_r^{(d)} \in [0, 1]$ the degree of discretization in stratum r relative to the maximum allowable by the observed data, that is,

$$\phi_r^{(d)} = \frac{K_r^{(d)}}{\sum_{i,j,k,s}^{(d)} \delta_{rijks}^{(d)}} \ ,$$

For large samples, it is very computationally inefficient to use $\phi^{(d)} = 1$, as evidenced in the results of Section 3.4.4, so for practical applications, where model selection and bootstrapped standard errors may be desired, a good choice of discretization level is critical. For simplicity, suppose that $\phi_r^{(d)} = \phi$ for $d \in \{0, 1\}$, $r = 1 \dots p$, although this not need to be the case: if there are many more events of one type, a fine level of discretization may be required for the rare process, while a coarse discretization may capture sufficient information about the frequent process.

For large samples, parameter estimates often do not change much beyond a sufficient discretization threshold. If multiple fits need to be computed for model

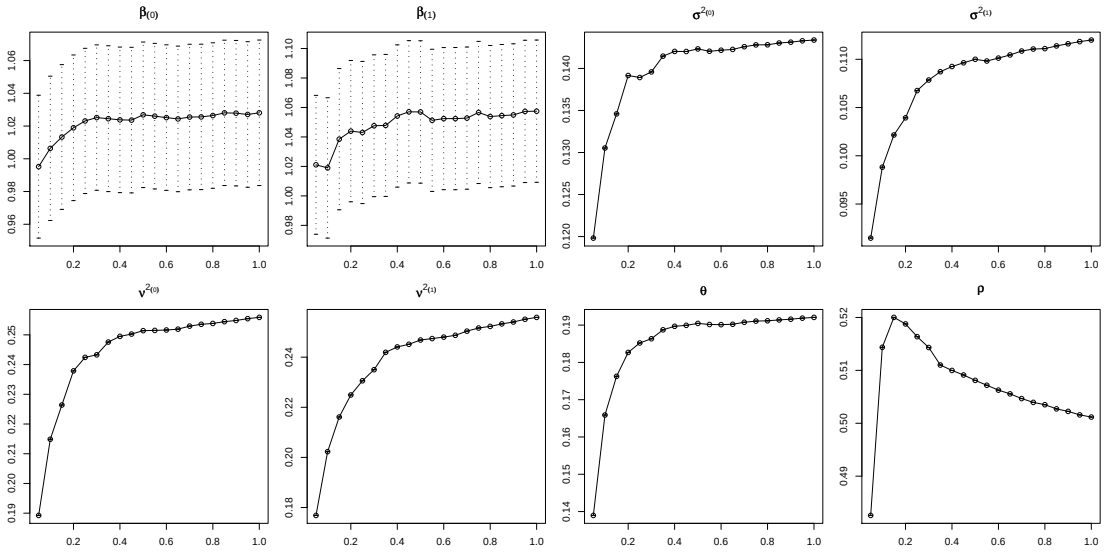

Figure 3.2: Parameter estimates and standard errors for covariates, for a simulated data set, plotted against the discretization parameter ϕ.

selection or bootstrapped standard errors, a discretization level close to the threshold may provide sufficiently accurate results while minimizing computation times.

As an example, consider a single data set generated following the simulation approach outlined in the print publication, with $m = 25$ clusters of $J_i = 25$ subjects each, according to setting (I). The resulting example set consists of 2273 events of type 0 and 1910 events of type 1. Parameter estimates for fitting the model at different discretization levels ϕ are plotted in fig. 3.2. The figures show a definite "elbow", and beyond $\phi = .4$, the parameter estimates do not change up to two significant digits.

The data analysis in Section 3.7 gives a practical application where coarser discretization allowed for time savings during model selection.

3.4.3 Setting discretization interval boundaries

For fixed $K_r^{(d)}$, breakpoints $a_{rs}^{(d)}$ must be chosen in a way that ensures that each interval contains at least one event. As before, denote $\phi_r^{(d)} = K_r^{(d)} \ / \ \sum_{i,j,k,s}^{(d)} \delta_{rijks}^{(d)}$, the level of discretization.

The simplest way to choose breakpoints automatically is as quantiles of the ordered event times in each stratum. That is, for $\phi = 1$, each event time constitutes a breakpoint, for $\phi = 0.5$, each interval contains two events, and for interim values of ϕ, quantiles can be interpolated in such a way that every interval contains either one or two events.

If events of equal length are desired, the following simple recursive procedure yields intervals of approximately equal length while satisfying the condition that each interval should contain an event. Suppose events of type d in cluster r occurring at sorted times $t_1 \dots t_n$, set $K = K_r^{(d)}$ $a_{r0}^{(d)} = 0$ and $a_{rK} = t_n$. Then,

1. If $K = 1$, terminate.
2. Propose a breakpoint $a = t_n \cdot \frac{K-1}{K}$
3. If $\sum_i I(t_i \geq a) \geq 1$ and $\sum_i I(t_i < a) \geq K - 1$, accept $a_{r(K-1)}^{(d)} = a$ and repeat the procedure for $K - 1$ breakpoints with only the times t_i such that $t_i < a$.
4. Otherwise, set $a_{r(K-1)}^{(d)}$ to the point nearest to a that satisfies the conditions in (3) and repeat.

For $\phi = 1$, this method simply yields the sample quantiles, but for very small values of ϕ the results can be quite different.

For the simulations in Section 3.A, quantiles of the ordered event times were used to define the breakpoints. Other simulation results indicate that using equal-length intervals does not significantly affect performance. For practical applications, either quantile or equal-length discretization intervals may be chosen depending on the shape of the baseline hazard.

3.4.4 Effect of discretization on computer time

For purposes of illustrating the effect of discretization on the computation time needed to fit the model, we generated simulated data sets of various sample sizes and fit them at different levels of discretization. Table 3.1 shows the time in rounded seconds it took to initialize, fit, and compute standard errors as a function of the number of clusters, cluster size, and discretization level. Computations were done on a 3.2GHz Pentium 4 processor with 1GB of memory, running R 2.6.0 on Ubuntu Linux 7.4. For certain large samples at fine discretization, the Godambe matrix was too large due to the large number of nuisance parameters, and it could not be inverted within a reasonable time. These cases are indicated in the table by a dash.

The table shows the advantages of using discretization for fitting large data sets using the proposed method. Firstly, the time to fit the model increases approximately linearly with the discretization parameter ϕ. In situations where a large number of models need to be fit, such as model selection or bootstrapping, this difference can be critical. Perhaps still more importantly, the time to compute standard errors involves the inversion of the Godambe matrix, and therefore grows at $O(n^3)$, where n is the total number of parameters, including the nuisance parameters $\alpha_{rs}^{(d)}$. For very large samples, time and computer memory limits may

Table 3.1: Computer time to fit simulated data generated under setting (I), for different sample sizes and discretization levels. For each cluster number m and cluster size J_i, the table contains the number of events in the generated data set $(M^{(0)}, M^{(1)})$. For each discretization level ϕ, three times are given: initialization time, EM algorithm fitting time, and standard error computation time. All times are in seconds. Dashes in the table indicate cases when the Godambe matrix was too large to invert numerically and standard errors could not be computed.

m	J_i	$M^{(0)}$	$M^{(1)}$	ϕ			
				0.1	0.25	0.5	1.0
10	10	322	357	0	0	0	0
				3	4	5	7
				0	0	0	1
	25	809	808	1	1	1	1
				4	6	8	14
				0	1	2	8
	50	1540	1739	3	3	3	3
				27	58	78	52
				1	7	18	106
	100	3458	3305	10	10	11	12
				45	91	146	283
				5	31	124	498
25	10	725	722	1	1	1	1
				9	12	16	25
				0	0	2	7
	25	2081	2072	5	5	5	5
				29	53	84	154
				2	8	32	128
	50	4532	4465	18	19	19	21
				123	141	212	456
				13	76	296	–
	100	7991	7068	57	58	60	65
				430	496	821	1825
				77	469	1269	–
50	10	1634	1733	3	3	3	4
				16	28	47	81
				1	5	21	81
	25	4420	4200	18	18	19	20
				141	209	301	617
				11	67	261	–
	50	8466	8617	65	67	70	76
				469	608	1074	2026
				86	520	–	–
	100	15953	15813	229	236	246	269
				735	1810	3400	6844
				523	–	–	–
100	10	3058	2903	11	11	12	12
				51	95	156	278
				5	27	104	590
	25	8663	8098	67	69	71	78
				244	530	925	1861
				78	474	–	–
	50	17351	15881	243	248	260	283
				779	1890	3450	6835
				557	–	–	–
	100	31222	32551	867	884	929	1011
				2724	7396	14025	28852
				7817	–	–	–

make it infeasible to compute standard errors if a fine level of discretization is used.

3.5 Extensions and modifications

In this section we consider several modifications and extensions to the model definition and fitting procedure described in Section 3.2. Specifically, Section 3.5.1 suggests a bias correction for the dispersion parameter estimates, Section 3.5.2 presents alternative marginal estimators for the dispersion parameters, and Section 3.5.3 discusses how the model would change if a different covariance structure were postulated for the frailties.

3.5.1 Bias corrections for frailty estimators

While the BLUP estimators in Section 3.2.3.1 are computationally straightforward, the estimators for the variance components of Section 3.2.3.2 are less so. The Pearson estimators on their own have a strong downward bias, and the bias-correction terms must be computed based on the previous estimates of the dispersion parameters. Therefore, the dispersion parameter estimates are prone to multiple sources of errors, most notably errors in the previous iteration's dispersion parameter estimates, which propagate into the frailty BLUP estimates and thence to the regression parameter estimates.

In order to correct for the downward bias in the dispersion parameter estimates, Ma (1999) suggested a degree-of-freedom correction, in which the estimators of

(3.19) are replaced by

$$\begin{aligned}
\hat{\sigma}^2_{(d)} &= \frac{1}{m-r}\sum_{i=1}^{m}\left\{(\hat{U}_i^{(d)}-1)^2 + b_i^{(d)}\right\}\ , \\
\hat{\nu}^2_{(d)} &= \frac{1}{m-r}\sum_{i=1}^{m}\frac{1}{J_i}\sum_{j=1}^{J_i}\left\{(\hat{U}_{ij}^{(d)}-\hat{U}_i^{(d)})^2 + b_{ij}^{(d)}\right\}\ , \\
\hat{\theta} &= \frac{1}{m-r}\sum_{i=1}^{m}\frac{1}{J_i}\sum_{j=1}^{J_i}\left\{(\hat{U}_{ij}^{(0)}-\hat{U}_i^{(0)})(\hat{U}_{ij}^{(1)}-\hat{U}_i^{(1)}) + b_{ij}^{(*)}\right\}\ ,
\end{aligned} \tag{3.56}$$

where r is the number of regression parameters that need to be estimated, that is, the length of $[\beta^{(0)}, \beta^{(1)}]$. Such a correction may be effective in reducing the downward bias of the dispersion parameter estimates for small samples, as suggested by the simulations in table 3.21.

3.5.2 Marginal dispersion parameter estimators

Under the auxiliary Poisson model, the marginal moments of the event indicators $\delta^{(d)}_{rijks}$ are known, as given in Section 3.3.1.1. This makes it possible to construct method of moments estimators for the dispersion parameters based on the known moments, conditional on the regression and baseline parameters, similar to those presented in Xue (1998). Given $\beta^{(d)}, \alpha^{(d)}_{rs}$, the following estimators for the variance components are unbiased:

$$\hat{\sigma}^2_{(d)} = \frac{\sum_{i=1}^{m} \sum_{(rjks,qbce):j\neq b} Y^{(d)}_{rijks} Y^{(d)}_{qibce} \left(\delta^{(d)}_{rijks} - \mu^{(d)}_{rijks}\right)\left(\delta^{(d)}_{qibce} - \mu^{(d)}_{qibce}\right)}{\sum_{i=1}^{m} \sum_{(rjks,qbce):j\neq b} Y^{(d)}_{rijks} Y^{(d)}_{qibce} \mu^{(d)}_{rijks} \mu^{(d)}_{ribce}}$$

$$\hat{\nu}^2_{(d)} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{J_i} \sum_{(rks,qce)} Y^{(d)}_{rijks} Y^{(d)}_{qijce} \left(\delta^{(d)}_{rijks} - \mu^{(d)}_{rijks}\right)\left(\delta^{(d)}_{qijce} - \mu^{(d)}_{qijce}\right) - \mu^{(d)}_{\cdot}}{\sum_{i=1}^{m} \sum_{j=1}^{J_i} \sum_{(rks,qce)} Y^{(d)}_{rijks} Y^{(d)}_{qijce} \mu^{(d)}_{rijks} \mu^{(d)}_{qijce}}$$

$$- \hat{\sigma}^2_{(d)}$$

$$\hat{\theta} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{J_i} \sum_{(rks,qce)} Y^{(0)}_{rijks} Y^{(1)}_{qijce} \left(\delta^{(0)}_{rijks} - \mu^{(0)}_{rijks}\right)\left(\delta^{(1)}_{qijce} - \mu^{(1)}_{qijce}\right)}{\sum_{i=1}^{m} \sum_{j=1}^{J_i} \sum_{(rks,qce)} Y^{(0)}_{rijks} Y^{(1)}_{qijce} \mu^{(0)}_{rijks} \mu^{(1)}_{qijce}}$$

These moment estimators are based on the covariance structure of Section 3.3.1.1. For $j \neq b$, the covariance structure implies that

$$\mathbb{E}\left[\left(\delta^{(d)}_{rijks} - \mu^{(d)}_{rijks}\right)\left(\delta^{(d)}_{qibce} - \mu^{(d)}_{qibce}\right)\right] = \text{Cov}\left(\delta^{(d)}_{rijks}, \delta^{(d)}_{qibce}\right) = \sigma^2_{(d)} \mu^{(d)}_{rijks} \mu^{(d)}_{qibce} ,$$

which justifies the estimator $\hat{\sigma}^2_{(d)}$. Similarly,

$$\mathbb{E}\left[\left(\delta^{(d)}_{rijks} - \mu^{(d)}_{rijks}\right)\left(\delta^{(d)}_{qijce} - \mu^{(d)}_{qijce}\right)\right]$$
$$= \text{Cov}\left(\delta^{(d)}_{rijks}, \delta^{(d)}_{qijce}\right)(\sigma^2_{(d)} + \nu^2_{(d)})\mu^{(d)}_{rijks}\mu^{(d)}_{qijce} + \mathbf{1}_{(rks,qce)}\mu^{(d)}_{rijks} ,$$

which means that the estimator

$$\hat{\nu}^2_{(d)} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{J_i} \sum_{(rks,qce)} Y^{(d)}_{rijks} Y^{(d)}_{qijce} \left(\delta^{(d)}_{rijks} - \mu^{(d)}_{rijks}\right)\left(\delta^{(d)}_{qijce} - \mu^{(d)}_{qijce}\right) - \mu^{(d)}_{\cdot}}{\sum_{i=1}^{m} \sum_{j=1}^{J_i} \sum_{(rks,qce)} Y^{(d)}_{rijks} Y^{(d)}_{qijce} \mu^{(d)}_{rijks} \mu^{(d)}_{qijce}}$$

is unbiased for $\sigma^2_{(d)} + \nu^2_{(d)}$, so that the estimator for $\nu^2_{(d)}$ is justified.

Simulations in table 3.18 show that replacing the Pearson-type estimators of Section 3.2.3.2 by the marginal estimators also results in parameter estimates that are slightly downward biased, but appear to be asymptotically consistent. The marginal estimators appear to perform slightly worse than the Pearson estimators, although the difference may be due to the simulation methodology.

3.5.3 Other frailty moment structures

The conditional moment structure for the subject-level frailties proposed in eqs. (3.1)–(3.4) is not the only possible reasonable covariance structure. If there is a compelling reason to require a frailty model in which the subject-level frailty covariance depends on cluster frailties, or a shared frailty model, minor changes to the BLUP estimators can easily accommodate these scenarios.

3.5.3.1 Subject covariance depending on cluster frailties

Suppose that in addition to (3.1), the subject-level frailties have moments

$$\begin{aligned}
\mathbb{E}\left[U_{ij}^{(d)} | U_{*}^{(d)} = u_{*}^{(d)}\right] &= u_i^{(d)} \\
\mathrm{Var}\left(U_{ij}^{(d)} | U_{*}^{(d)} = u_{*}^{(d)}\right) &= u_i \nu_{(d)}^2 \\
\mathrm{Cov}\left(U_{ij}^{(0)}, U_{ij}^{(1)} | U_{*}^{(*)} = u_{*}^{(*)}\right) &= \rho \nu_{(0)} \nu_{(1)} u_i^{(0)} u_i^{(1)} \,,
\end{aligned}$$

so that ρ denotes the conditional correlation of the frailties. Computations analogous to Section 3.3.1 give BLUPs as in (3.15) and (3.16), but with

$$w_{ij} = \left(\prod_{d \in \{0,1\}} (1 + \nu_{(d)}^2 (\sigma_{(d)}^2 + 1) \mu_{ij.}^{(d)}) - \rho^2 \prod_{d \in \{0,1\}} \nu_{(d)}^2 \mu_{ij.}^{(d)} \right)^{-1} ,$$

and

$$\begin{aligned}
P_{ij}^{(d)} &= \delta_{ij.}^{(d)} \cdot w_{ij} \left(1 + \nu_{(1-d)}^2 (\sigma_{(1-d)}^2 + 1) \mu_{ij.}^{(1-d)}\right) \quad Q_{ij}^{(d)} = -\delta_{ij.}^{(d)} \cdot w_{ij} \rho \nu_{(d)} \nu_{(1-d)} \mu_{ij.}^{(1-d)} \\
p_{ij}^{(d)} &= \mu_{ij.}^{(d)} \cdot w_{ij} \left(1 + \nu_{(1-d)}^2 (\sigma_{(1-d)}^2 + 1) \mu_{ij.}^{(1-d)}\right) \quad q_{ij}^{(d)} = -\mu_{ij.}^{(d)} \cdot w_{ij} \rho \nu_{(d)} \nu_{(1-d)} \mu_{ij.}^{(1-d)}
\end{aligned}$$

replacing (3.17) and (3.18) respectively.

Bias corrections and standard error estimators can be constructed analogously.

3.5.3.2 Shared frailties

Supposing that $U_i^{(0)} = U_i^{(1)} = U_i$ and $U_{ij}^{(0)} = U_{ij}^{(1)} = U_{ij}$, with $\mathrm{Var}\,(U_i) = \sigma^2$ and $\mathrm{Var}\,(U_{ij}|U_i) = \nu^2$ leads to a shared frailty model structurally similar to that of Ma (1999). The BLUPs are then given by

$$\hat{U}_i = \frac{1 + \sigma^2 P_i}{1 + \sigma^2 p_i}\,, \qquad \hat{U}_{ij} = (1 - \nu^2 p_{ij})\hat{U}_i + \nu^2 P_{ij}$$

where $p_i = \sum_{i=1}^{m} p_{ij}$ and $P_i = \sum_{i=1}^{m} P_{ij}$, and

$$P_{ij} = \frac{\delta_{ij.}^{(.)}}{1 + \nu^2 \mu_{ij.}^{(.)}}\,, \qquad p_{ij} = \frac{\mu_{ij.}^{(.)}}{1 + \nu^2 \mu_{ij.}^{(.)}}\,.$$

Bias correction terms and standard error estimators are analogous to those of Ma (1999).

3.6 Simulation studies

We have implemented the methodology described in Section 3.2 in the `R` package `blupsurv`. Because the asymptotic properties of the proposed methodology are not rigorously established, several simulation studies were conducted in order to demonstrate its performance in a variety of settings. We present results for typical settings here, including simulations covering alternative settings and extensions in Section 3.A. Section 3.6.1 presents the simulation methodology, and Section 3.6.2 summarizes the results.

Table 3.2: True regression and dispersion parameter values used to generate simulated samples.

Setting	$\beta^{(0)}$	$\beta^{(1)}$	$\sigma^2_{(0)}$	$\sigma^2_{(1)}$	$\nu^2_{(0)}$	$\nu^2_{(1)}$	θ
(I)	1	1	0.25	0.25	0.25	0.25	0.125
(II)	1	2	0.25	0.25	0.25	0.25	0.125
(III)	1	1	0.25	0.5	0.25	0.5	0.125

3.6.1 Simulation methodology

Simulations were conducted to investigate performance as a function of the number of clusters, the number of subjects per cluster, and the degree of discretization. Each simulation consisted of generating many simulated datasets from a specified parametrized distribution and using the `blupsurv` package to estimate regression and dispersion parameters.

Three settings are considered, differing in the "true" regression and dispersion parameters used to generate the samples. Parameter values for all simulations reported in this paper are summarized in Table 3.2. In setting (I), both processes are generated using the same regression and dispersion parameters, with settings (II) and (III) respectively allowing the regression parameters and dispersion parameters to be different.

Within each setting, we considered four sample sizes, setting the number of clusters set to either $m = 10$ or $m = 25$, and the cluster size to either $J_i = 5$ or $J_i = 25$. For each setting and sample size, we conducted 1000 replications.

Each replication consisted of generating simulated frailties $(U_i^{(d)}, U_{ij}^{(d)};\ i = 1 \ldots m, j = 1 \ldots J_i, d \in \{0, 1\})$ from a hierarchical log-Normal distribution with the moment structure specified in (3.1) through (3.4), with the appropriate dispersion parameters for that setting. A single covariate was generated for each

subject via $Z \sim N(0, .5)$. Recurrent event gap times were generated using single-stratum Weibull baseline hazards with scale parameters $\lambda_0^{(0)} = \lambda_0^{(1)} = 10$ and shape parameters $\eta_0^{(0)} = \eta^{(1)} = 1.8$, so that the baseline hazards were given by $\lambda_{01}^{(d)}(t) = \lambda_0^{(d)}\eta_0^{(d)}t^{\eta_0^{(d)}-1}$. Censoring times were independently generated from a Weibull hazard with parameters $\lambda_c = 1, \eta_c = 1.8$. These parameters were chosen so that a subject under setting (I) experienced four events of each type (i.e., on average).

Each dataset was fit at four levels of discretization, parametrized by a parameter $\phi = (0.1, 0.25, 0.5, 1.0)$ specifying the ratio of the number of discretization intervals used to the maximum permissible by the data, that is,

$$K_r^{(d)} = \phi \cdot \sum_{i,j,k,s} \delta_{rijks}^{(d)} \ .$$

For simplicity, the level of discretization was applied in equal proportion to both event types for the simulation; however, this is not necessary in applications.

3.6.2 Simulation results

Tables 3.3, 3.4 and 3.5 respectively contain the results of simulations conducted under settings (I)-(III). In each table, Panel A contains the bias of the parameter estimates for various sample sizes and discretization levels and Panel B gives the estimated standard errors and 95% confidence interval coverage rates for the regression parameter estimates. The corresponding variances and mean squared errors are provided in Section 3.A.

Panel A of Table 3.3 shows that under setting (I), regression parameters are well-estimated for large sample sizes and sufficiently fine discretization, and dispersion parameter estimates have small negative biases. In general, the biases

Table 3.3: Bias and standard error of parameter estimates from 1000 simulations of two recurrent event processes under setting (I). CSE is mean computed standard error, ESE is empirical standard error, 95%CP is the coverage rate of 95% confidence intervals derived from CSE.

Panel A: Bias of regression coefficients and dispersion parameters (Setting I)

m	J_i	ϕ	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\sigma}^2_{(0)}$	$\hat{\sigma}^2_{(1)}$	$\hat{\nu}^2_{(0)}$	$\hat{\nu}^2_{(1)}$	$\hat{\theta}$	$\hat{\rho}$
10	5	.1	−0.076	−0.077	−0.130	−0.135	−0.146	−0.148	−0.073	0.001
		.25	−0.033	−0.024	−0.076	−0.084	−0.089	−0.099	−0.047	0.008
		.5	−0.022	−0.007	−0.079	−0.064	−0.069	−0.050	−0.032	0.029
		1	−0.000	−0.010	−0.072	−0.061	−0.062	−0.065	−0.028	0.027
	25	.1	−0.024	−0.024	−0.056	−0.057	−0.065	−0.066	−0.022	0.046
		.25	−0.005	−0.007	−0.043	−0.044	−0.037	−0.039	−0.008	0.050
		.5	−0.005	0.001	−0.054	−0.056	−0.043	−0.038	−0.010	0.051
		1	−0.001	0.001	−0.051	−0.063	−0.056	−0.057	−0.017	0.044
25	5	.1	−0.046	−0.046	−0.066	−0.069	−0.102	−0.111	−0.046	0.002
		.25	−0.023	−0.023	−0.039	−0.045	−0.071	−0.070	−0.031	0.005
		.5	−0.014	−0.008	−0.051	−0.050	−0.062	−0.057	−0.022	0.024
		1	−0.001	−0.010	−0.042	−0.044	−0.057	−0.058	−0.024	0.012
	25	.1	−0.008	−0.013	−0.024	−0.024	−0.047	−0.051	−0.015	0.019
		.25	−0.001	−0.006	−0.021	−0.022	−0.040	−0.045	−0.010	0.023
		.5	−0.002	0.000	−0.030	−0.025	−0.047	−0.051	−0.014	0.021
		1	−0.003	−0.000	−0.032	−0.033	−0.061	−0.061	−0.020	0.017

Panel B: Standard error of regression coefficients (Setting I)

			$\hat{\beta}^{(0)}$			$\hat{\beta}^{(1)}$		
m	J_i	ϕ	CSE	ESE	95%CP	CSE	ESE	95%CP
10	5	.1	0.211	0.243	0.889	0.210	0.248	0.888
		.25	0.227	0.263	0.909	0.226	0.252	0.922
		.5	0.233	0.265	0.908	0.236	0.273	0.903
		1	0.231	0.280	0.901	0.231	0.269	0.910
	25	.1	0.098	0.105	0.923	0.098	0.108	0.914
		.25	0.100	0.108	0.931	0.100	0.104	0.928
		.5	0.100	0.107	0.934	0.101	0.103	0.941
		1	0.098	0.114	0.899	0.098	0.110	0.912
25	5	.1	0.140	0.159	0.892	0.139	0.151	0.909
		.25	0.146	0.167	0.910	0.146	0.160	0.918
		.5	0.146	0.164	0.931	0.147	0.162	0.923
		1	0.147	0.165	0.920	0.147	0.165	0.918
	25	.1	0.063	0.069	0.925	0.063	0.068	0.932
		.25	0.063	0.068	0.936	0.063	0.071	0.924
		.5	0.062	0.069	0.924	0.062	0.070	0.917
		1	0.061	0.064	0.935	0.061	0.066	0.932

Table 3.4: Bias and standard error of parameter estimates from 1000 simulations of two recurrent event processes under setting (II), which differs from setting (I) only in that $\beta^{(1)} = 2$. CSE is mean computed standard error, ESE is empirical standard error, 95%CP is the coverage rate of 95% confidence intervals derived from CSE.

Panel A: Bias of regression coefficients and dispersion parameters (Setting II)

m	J_i	ϕ	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\sigma}^2_{(0)}$	$\hat{\sigma}^2_{(1)}$	$\hat{\nu}^2_{(0)}$	$\hat{\nu}^2_{(1)}$	$\hat{\theta}$	$\hat{\rho}$
10	5	.1	−0.069	−0.140	−0.127	−0.136	−0.146	−0.149	−0.072	0.000
		.25	−0.034	−0.065	−0.089	−0.087	−0.081	−0.089	−0.039	0.036
		.5	−0.011	−0.033	−0.080	−0.074	−0.062	−0.060	−0.029	0.038
		1	0.007	−0.004	−0.074	−0.070	−0.068	−0.058	−0.035	0.018
	25	.1	−0.029	−0.045	−0.059	−0.067	−0.070	−0.075	−0.024	0.050
		.25	−0.002	−0.018	−0.048	−0.051	−0.038	−0.049	−0.012	0.049
		.5	−0.010	−0.005	−0.056	−0.055	−0.045	−0.046	−0.011	0.050
		1	−0.004	−0.007	−0.052	−0.070	−0.062	−0.062	−0.020	0.043
25	5	.1	−0.052	−0.089	−0.073	−0.080	−0.102	−0.097	−0.042	0.016
		.25	−0.016	−0.032	−0.053	−0.045	−0.069	−0.074	−0.029	0.014
		.5	−0.013	−0.022	−0.043	−0.049	−0.056	−0.058	−0.022	0.020
		1	−0.005	−0.002	−0.045	−0.049	−0.056	−0.062	−0.025	0.013
	25	.1	−0.010	−0.022	−0.017	−0.031	−0.043	−0.057	−0.015	0.019
		.25	−0.003	−0.004	−0.025	−0.025	−0.043	−0.046	−0.012	0.023
		.5	−0.004	−0.007	−0.024	−0.025	−0.049	−0.050	−0.014	0.018
		1	−0.002	−0.003	−0.039	−0.037	−0.063	−0.060	−0.020	0.020

Panel B: Standard error of regression coefficients (Setting II)

			$\hat{\beta}^{(0)}$			$\hat{\beta}^{(1)}$		
m	J_i	ϕ	CSE	ESE	95%CP	CSE	ESE	95%CP
10	5	.1	0.212	0.250	0.879	0.233	0.289	0.837
		.25	0.227	0.257	0.908	0.249	0.290	0.896
		.5	0.233	0.269	0.906	0.256	0.287	0.902
		1	0.232	0.265	0.926	0.256	0.304	0.902
	25	.1	0.097	0.106	0.915	0.106	0.118	0.886
		.25	0.101	0.111	0.927	0.109	0.125	0.908
		.5	0.100	0.111	0.918	0.109	0.123	0.909
		1	0.097	0.113	0.897	0.106	0.124	0.906
25	5	.1	0.141	0.153	0.901	0.155	0.170	0.869
		.25	0.146	0.162	0.924	0.159	0.180	0.913
		.5	0.148	0.163	0.924	0.161	0.176	0.936
		1	0.147	0.161	0.916	0.160	0.190	0.895
	25	.1	0.063	0.070	0.926	0.068	0.075	0.902
		.25	0.063	0.066	0.932	0.069	0.079	0.917
		.5	0.062	0.068	0.937	0.068	0.079	0.905
		1	0.061	0.069	0.907	0.067	0.077	0.907

Table 3.5: Bias and standard error of parameter estimates from 1000 simulations of two recurrent event processes under setting (III), which differs from setting (I) only in that $\sigma^2_{(1)} = \nu^2_{(1)} = .5$. CSE is mean computed standard error, ESE is empirical standard error, 95%CP is the coverage rate of 95% confidence intervals derived from CSE.

Panel A: Bias of regression coefficients and dispersion parameters (Setting III)

m	J_i	ϕ	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\sigma}^2_{(0)}$	$\hat{\sigma}^2_{(1)}$	$\hat{\nu}^2_{(0)}$	$\hat{\nu}^2_{(1)}$	$\hat{\theta}$	$\hat{\rho}$
10	5	.1	−0.082	−0.125	−0.124	−0.287	−0.151	−0.325	−0.065	0.036
		.25	−0.014	−0.068	−0.080	−0.164	−0.084	−0.147	−0.025	0.057
		.5	−0.027	−0.035	−0.066	−0.138	−0.069	−0.136	−0.024	0.046
		1	0.011	−0.028	−0.060	−0.153	−0.065	−0.129	−0.015	0.057
	25	.1	−0.035	−0.032	−0.046	−0.068	−0.070	−0.029	−0.001	0.062
		.25	−0.014	−0.007	−0.041	−0.172	−0.043	−0.042	0.007	0.074
		.5	−0.010	−0.007	−0.046	−0.178	−0.043	−0.040	0.004	0.067
		1	−0.006	−0.006	−0.052	−0.180	−0.063	−0.076	−0.004	0.061
25	5	.1	−0.044	−0.066	−0.065	−0.129	−0.108	−0.194	−0.032	0.034
		.25	−0.028	−0.034	−0.028	−0.064	−0.065	−0.121	−0.012	0.033
		.5	−0.015	−0.018	−0.033	−0.057	−0.062	−0.095	−0.014	0.026
		1	−0.018	−0.019	−0.035	−0.076	−0.060	−0.111	−0.012	0.033
	25	.1	−0.012	−0.005	−0.015	−0.056	−0.042	0.035	0.012	0.041
		.25	−0.005	0.009	−0.009	−0.086	−0.045	0.012	0.007	0.037
		.5	−0.007	0.006	−0.016	−0.084	−0.049	−0.022	0.002	0.035
		1	−0.004	0.004	−0.025	−0.082	−0.063	−0.083	−0.006	0.031

Panel B: Standard error of regression coefficients (Setting III)

			$\hat{\beta}^{(0)}$			$\hat{\beta}^{(1)}$		
m	J_i	ϕ	CSE	ESE	95%CP	CSE	ESE	95%CP
10	5	.1	0.211	0.252	0.875	0.231	0.285	0.829
		.25	0.231	0.273	0.898	0.268	0.299	0.900
		.5	0.234	0.268	0.905	0.272	0.316	0.893
		1	0.234	0.262	0.920	0.271	0.314	0.906
	25	.1	0.098	0.108	0.899	0.124	0.132	0.919
		.25	0.101	0.107	0.925	0.124	0.130	0.929
		.5	0.101	0.107	0.913	0.123	0.133	0.930
		1	0.098	0.109	0.924	0.119	0.133	0.925
25	5	.1	0.140	0.162	0.894	0.164	0.187	0.886
		.25	0.147	0.162	0.912	0.174	0.185	0.912
		.5	0.147	0.169	0.908	0.176	0.192	0.911
		1	0.147	0.164	0.918	0.173	0.197	0.915
	25	.1	0.063	0.068	0.920	0.082	0.086	0.941
		.25	0.063	0.070	0.911	0.080	0.082	0.953
		.5	0.062	0.068	0.917	0.078	0.083	0.937
		1	0.061	0.067	0.924	0.074	0.081	0.934

reported here compare favorably to those reported in Ma and Jørgensen (2007) and Ha and Lee (2005) for the univariate case. Increasing either (or both) the number of clusters or the cluster size leads to improved dispersion parameter estimates; as might be expected, the bias of $\sigma^2_{(d)}$ is primarily affected by the number of clusters, whereas cluster size tends to be of greater importance when estimating the subject-level dispersion parameters (i.e., $\nu^2_{(d)}, \theta$).

Of some interest in these tables is the effect of discretization: for small samples, successively finer discretization produces much improved results. However, with larger sample sizes, the results additionally suggest that such fine discretization is not necessary. For the largest samples with $m = 25, J_i = 25$, the estimation bias does not decrease appreciably beyond $\phi = .25$, and for samples of intermediate size, the performance gain with increased discretization is quite modest. The benefit of proper discretization is readily apparent: for large samples, good results can be obtained with relatively little computational effort by a judicious choice of discretization.

In order to investigate the finite-sample performance of the covariance matrix estimate proposed in Section 3.2.4, Panel B of Table 3.3 provides standard errors and 95% confidence interval coverage rates for the regression parameter estimates in setting (I). As commented earlier, we observe a slight underestimation of standard error, the degree of which appears similar to that reported in Ma and Jørgensen (2007). We additionally observe that the extent of underestimation is more severe in small samples with coarse discretization, improving with finer discretization and both increasing m and $J_1 \ldots J_m$. The fact that the standard error estimates degrade as the level of discretization increases is consistent with our earlier conjecture that the basic score equation (3.23) may not be exactly unbiased

in finite samples. In any event, this systematic underestimation of standard errors should be taken into account when interpreting the magnitudes of p-values.

Tables 3.4 and 3.5 respectively contain analogous results for settings (II) and (III). Results indicate that while varying regression and dispersion parameters does affect the absolute bias, the percentage bias remains approximately constant. There is also no clear indication that varying either the regression or dispersion parameters signficantly impacts the relative performance of the covariance and correlation parameter estimates. Further simulation results for time-dependent covariates, alternative dispersion parameter estimators, and other methods of bias corrections may be found in Appendix 3.A.

3.7 Data Examples

In this section, we apply the proposed methodology to data collected from two randomized clinical trials. In Section 3.7.1, we reanalyze data originally collected as part of the Nutritional Prevention of Cancer Trial conducted by the Arizona Cancer Center between 1985 and 1996 (Clark et al., 1996; Duffield-Lillico et al., 2002). This study was designed to evaluate the effectiveness of selenium supplementation on prevention of nonmelanoma skin cancer, defined by the occurrence of basal or squamous cell carcinomas of the skin. In Section 3.7.2, we reanalyze data from a randomized double-blind study of pulmonary exacerbations in cystic fibrosis patients (Fuchs et al., 1994), where patients were treated with aerosolized recombinant human deoxyribonuclease (rhDNase) in the hope of reducing the frequency and length of exacerbation episodes. The bivariate process of interest here is created by the alternating sequence of times "between" and "within" exacer-

bation episodes. In both sections, our data analyses utilize $\phi = 1$ (i.e., no time discretization).

3.7.1 Effects of selenium supplementation on skin cancer

We summarize the study's methods and findings in Section 3.7.1.1, provide details on the application of our methods in Section 3.7.1.2, and summarize the results in Section 3.7.1.3.

3.7.1.1 Study methods and findings

The design and methods of the Nutritional Prevention of Cancer (NPC) trial are described in detail by Clark et al. (1996) and Duffield-Lillico et al. (2002); a brief summary is provided here for completeness. The NPC trial was a double-blind controlled study that followed a cohort of 1312 patients in seven dermatology clinics throughout the United States. Treatment consisted of a $0.5g$ tablet containing $200\mu g$ of selenium for the treatment group. Patients were initially evaluated on sun exposure and sensitivity, as well as prior BCC (Basal Cell Carcinoma) and SCC (Squamous Cell Carcinoma) history, and scheduled to return to the clinic in six month intervals. New BCC and SCC occurrences could be diagnosed by the patients' own dermatologists, but were also confirmed by biopsy at each clinic visit. At every visit, plasma selenium levels were measured in the laboratory for each patient.

Patient data gathered consisted of plasma selenium level (baseline level and laboratory measurements at each clinic visit), age, gender, height, weight, BMI, smoking status (current or ex-smoker, number of daily cigarettes, number of years

Table 3.6: Number of observed events, number of subjects affected by events, and incidence, by cancer type and treatment group. Total affected subjects do not sum because multiple types of cancer are counted only once.

	Placebo			Treatment			Total		
	Events	Subj.	Inc.	Events	Subj.	Inc.	Events	Subj.	Inc.
BCC	1263	370	0.256	1503	399	0.301	2766	769	0.279
SCC	479	192	0.097	568	246	0.114	1047	438	0.106
Total	1742	413	0.354	2071	462	0.415	3813	875	0.385

of smoking), alcohol consumption (number of weekly drinking days, number of drinks per day), a sun damage index, fasting status, use of vitamin supplements, use of sunscreen (always, sometimes, never), number of years spent on a farm, hair and eye color, and number of BCC, SCC and AK (Actinic Keratosis) events prior to randomization. Of these variables, only plasma selenium varies with time (i.e., a time-dependent covariate). Descriptive statistics for the outcomes and covariates are given in Table 3.6 and Table 3.7, respectively.

Clark et al. (1996) analyzed the data using Kaplan-Meier estimates and log-rank tests for the effect of treatment on BCC and SCC incidence. These analyses, done separately by cancer type, used the time to the first post-randomization cancer as the primary outcome variable and found a nonsignificant increase in the incidence of both cancer types as a result of treatment. Using the fully parametric mixed nonhomogenous Poisson process model described in Abu-Libdeh et al. (1990), Clark et al. (1996) also found a statistically insignificant increased risk due to treatment for the recurrent BCC and SCC outcomes.

3.7.1.2 Data analysis methodology

Plasma selenium levels were measured at baseline (i.e., study entry) and also post-randomization. However, because post-randomization levels of plasma selenium

Table 3.7: Basic descriptive statistics of observed covariates, for patients in placebo and treatment groups, or who suffered at least one BCC or SCC respectively.

	Placebo		Treatment		BCC		SCC	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Baseline selenium	114.01	21.48	114.38	22.60	114.72	22.65	112.80	22.05
Average selenium	113.99	15.97	168.78	43.72	143.12	43.02	142.33	42.05
Previous BCCs	1.24	1.56	1.42	1.96	1.70	1.95	1.21	1.62
Previous SCCs	0.58	0.81	0.57	0.79	0.47	0.77	0.82	0.95
Gender (% M)	0.75	0.43	0.74	0.44	0.80	0.40	0.86	0.35
Age	63.04	9.91	63.39	10.16	63.43	9.61	65.16	7.94
Height (m)	1.74	0.09	1.74	0.15	1.74	0.09	1.75	0.08
Weight (kg)	77.37	14.65	77.45	14.68	78.23	14.24	79.42	13.88
BMI	25.49	4.12	25.61	3.86	25.60	3.92	25.67	3.76
Smoking (current)	0.30	0.46	0.27	0.44	0.26	0.44	0.29	0.46
Smoking (ex)	0.40	0.49	0.40	0.49	0.42	0.49	0.45	0.50
Smoking (cigs/day)	17.24	17.03	16.42	17.63	16.92	17.56	18.86	17.64
Smoking (years)	31.62	22.68	29.75	23.09	31.14	23.05	35.04	22.11
Drinking (days/week)	1.24	2.24	1.47	2.43	1.51	2.44	1.55	2.52
Drinking (no. of drinks)	0.79	1.33	0.92	1.71	0.93	1.62	0.85	1.32
Farming (years)	7.21	8.10	7.76	8.11	7.03	7.98	8.30	8.24
Vitamins	0.39	0.49	0.36	0.48	0.39	0.49	0.35	0.48
Sun damage index	4.73	1.65	4.81	1.67	4.79	1.65	5.34	1.50
Sunscreen (sometimes)	0.43	0.50	0.44	0.50	0.45	0.50	0.45	0.50
Sunscreen (never)	0.28	0.45	0.28	0.45	0.31	0.46	0.31	0.46

are directly influenced by treatment and this variable represents the hypothesized mechanism by which treatment affects skin cancer incidence, it is inappropriate to include both the treatment group indicator and post-randomization plasma selenium levels as possible covariates when evaluating the treatment effect.

Treating the seven clinics as a cluster-level random effect, an analysis of the treatment effect using the proposed model continues to suggest a detrimental impact of treatment on both BCC recurrence ($p = 0.0265$) and SCC recurrence ($p = 0.0566$). We therefore consider two distinct post-hoc analyses. In the first analysis, we utilize only the baseline covariates available for each patient, including baseline selenium levels; we do not include plasma selenium levels measured post-randomization as a predictor variable. In the second analysis, we include the time-dependent plasma selenium level, but not the treatment indicator, thereby intending to evaluate the effect of changes in the longitudinal plasma selenium level on the recurrent event processes.

Extensive exploratory data analysis and conversations with the lead study physicians at the Roswell Park Cancer Institute led us to consider a particular subset of the baseline covariates in Table 3.7 for subsequent analysis. Missing values ($< 1\%$) were imputed with the corresponding median value. As indicated earlier, the seven clinics were treated as cluster-level random effects; no stratification was used. Patients whose blood was drawn more than four days after randomization were also excluded from the analysis, leaving a cohort of 1250 patients; see Duffield-Lillico et al. (2002) for further discussion. Further details on the variables selected for this analysis and the resulting model fit may be found in Section 3.7.1.3.

For the analysis of post-randomization plasma selenium levels, we considered

Table 3.8: Regression and dispersion parameter estimates for the bivariate clustered frailty model fit to the Nutritional Prevention of Cancer trial data, with baseline covariates only.

Regression parameter estimates							
		BCC			SCC		
		Coeff	Std	Pval	Coeff	Std	Pval
	Treatment	0.1586	0.0707	0.0125	0.2148	0.1145	0.0304
	Baseline Se < Median	−0.1597	0.0724	0.0137	0.0435	0.1155	0.3531
History:	No BCC or SCC	−0.9776	0.1578	<1e-04	0.0807	0.2535	0.3751
History:	SCC only	−1.1037	0.0929	<1e-04	0.9109	0.1388	<1e-04
History:	Both BCC and SCC	0.0958	0.0967	0.1609	1.0171	0.1623	<1e-04
AK History:	AK > 2	0.2644	0.0991	0.0038	0.6474	0.1443	<1e-04
Age/Gender:	Young Male	0.3532	0.1166	0.0012	0.8977	0.2204	<1e-04
Age/Gender:	Older Female	−0.1484	0.1519	0.1643	0.7119	0.2634	0.0034
Age/Gender:	Older Male	0.4178	0.1181	0.0002	1.3192	0.2177	<1e-04
	Drink Days > 2	0.1637	0.0863	0.0288	0.1038	0.1405	0.2300
	Sundamage > 5	0.0726	0.0804	0.1835	0.5844	0.1202	<1e-04

Dispersion parameter Estimates	
Cluster frailty dispersion (BCC)	0.0087
Cluster frailty dispersion (SCC)	0.0014
Subject frailty dispersion (BCC)	0.9332
Subject frailty dispersion (SCC)	2.3163
Subject frailty covariance	0.3438
Subject frailty correlation	0.2327

a modified version of the model used to conduct our baseline analysis. The time-dependent covariate capturing selenium was created by subtracting and scaling by each patient's initial selenium level; thus, it represents the percentage above or below the patient's baseline selenium level at each time. The resulting model was fit as described in Section 3.2.5.

3.7.1.3 Data analysis results

The baseline covariates utilized in our first analysis are defined as follows:

- **Treatment:** Takes value 1 if the patient was a member of the treatment group.

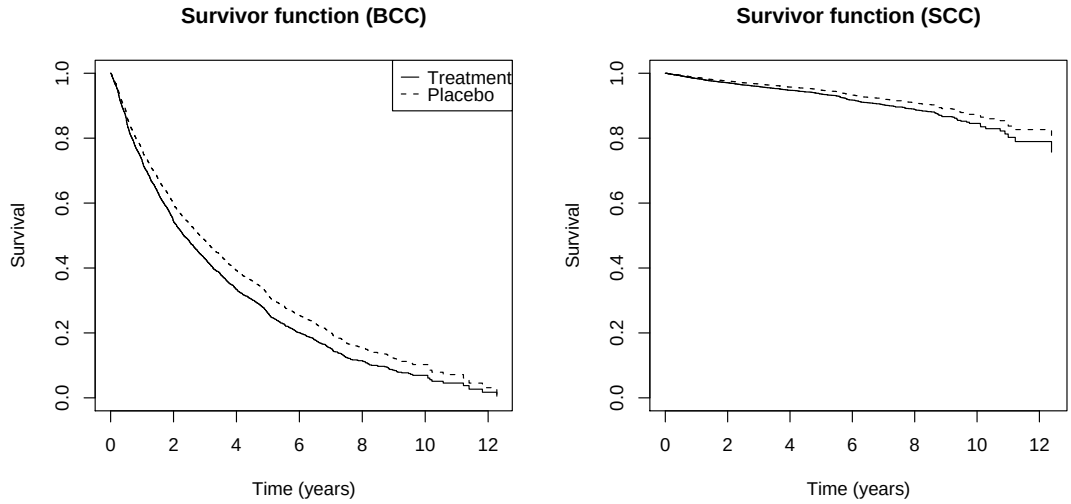

Figure 3.3: Estimated baseline survivor functions for BCC and SCC gap times, and survivor functions adjusted for the treatment effect.

- **Baseline Se $<$ Median:** Takes value 1 if the patient's baseline selenium level was below the median baseline level.

- **History:** Nearly 70% of patients in the trial presented with either: a prior history of BCCs but no SCC; or, a prior history of SCCs but no BCC. A patient's BCC and SCC history before randomization is described by three indicator variables. Patients who have suffered at least one BCC, but no SCCs (the most common case) serve as the baseline group. The remaining indicators correspond to patients with no history of either skin cancer type; a history of both cancer types; or a history of SCC only.

- **AK History:** Takes value 1 if a patient has experienced more than two AK events prior to randomization.

- **Age/Gender:** Patients are grouped by age and gender indicators. A patient is considered "older" if their age is above the global median (65). Young females serve as the baseline group.

- **Drink Days > 2:** Takes value 1 if the patient drinks more than two days per week (i.e., above the 75^{th} percentile)

- **Sundamage > 5** Takes value 1 if the patient's sun damage index is above the modal value of 5 (higher values indicate greater damage).

The estimated model coefficients and frailty dispersion parameters are summarized in Table 3.8. The results indicate that treatment is associated with an increase in both BCC and SCC incidence. There are matching adverse effects of high baseline selenium on both BCC and SCC, the latter being statistically insignificant. The estimated effects of skin cancer history suggest that patients who have experienced a given type of skin cancer are at particularly high risk for further events of the same type. A history of AK occurrences is highly predictive for SCC risk, but less so for BCC. Regular drinking is related to increased risk of BCC but not SCC, whereas sun damage significantly affects SCC risk only. It should be noted that standard errors were underestimated slightly in simulation, so these results may overstate statistical significance.

The estimated dispersion parameters indicate a strong subject-level frailty effect, but very small cluster-level variability for both BCC and SCC processes. In other words, the heterogeneity unaccounted for by the covariates may be substantially larger within clusters than between clusters. The covariance estimate indicates positive correlation between the frailties for the two processes; that is, higher event rates of one skin cancer type tend to occur with higher events rates of the other.

Figure 3.3 shows the estimated baseline survivor functions for the gap times between BCC and SCC events. The baseline refers to a female subject under 65

who has above-median baseline selenium, has suffered at least one BCC event but no SCC or AKs, drinks less than twice weekly and has low sun damage.

Remark: In Abu-Libdeh et al. (1990), an interim analysis of the NPC trial data is presented for the purposes of illustration. In that analysis, the treatment indicator was intentionally randomized in order to avoid influencing the study outcome; consequently, the results of analyses summarized there are not directly comparable to those summarized in this paper. In addition, the model employed by Abu-Libdeh et al. (1990) is quite different; in addition to being fully parametric, for example, covariate effects are also assumed to have a common impact on the intensity of both skin cancer types. The proposed model does not enforce such requirements and, as suggested by the results in Table 7, the latter assumption does not appear to be reasonable for certain covariates.

Next, we consider the effect of time-varying plasma selenium on the BCC and SCC processes. Because plasma selenium is the mechanism by which treatment is expected to affect the processes, the treatment group indicator was left out of the analysis. The covariate labeled "Selenium (% over baseline)" is defined as the time-varying percentage by which a patient's current selenium level exceeds the patient's selenium level at study entry. All other covariates included in the prior analysis were also included here; however, the dichotomized baseline selenium variable was replaced by its continuous counterpart. The results of this analysis are summarized in Table 3.9. The estimated effect of the continuous baseline selenim variable is consistent with that suggested in Table 3.8. It is also observed that the impact of rising selenium continues to be detrimental to both BCC and SCC recurrence; however, neither effect is statistically significant. The remaining covariate effects (parameter estimates, standard errors, p-values) are qualitatively

Table 3.9: Regression and dispersion parameter estimates for the bivariate clustered frailty model fit to the Nutritional Prevention of Cancer trial data, with time-dependent plasma selenium.

Regression parameter estimates							
		BCC			SCC		
		Coeff	Std	Pval	Coeff	Std	Pval
	Selenium (% over baseline)	0.0006	0.0006	0.1359	0.0011	0.0009	0.1188
	Baseline Selenium	0.0036	0.0016	0.0109	−0.0035	0.0026	0.0863
History:	No BCC or SCC	−0.9639	0.1486	<1e-04	0.0801	0.2409	0.3697
History:	SCC only	−1.0921	0.0880	<1e-04	0.8630	0.1316	<1e-04
History:	Both BCC and SCC	0.0913	0.0898	0.1546	0.9536	0.1534	<1e-04
AK History:	AK > 2	0.2743	0.0932	0.0016	0.6529	0.1357	<1e-04
Age/Gender:	Young Male	0.3415	0.1094	0.0009	0.9236	0.2116	<1e-04
Age/Gender:	Older Female	−0.1564	0.1422	0.1356	0.6498	0.2523	0.0050
Age/Gender:	Older Male	0.4072	0.1109	0.0001	1.2936	0.2087	<1e-04
	Drink Days > 2	0.1705	0.0802	0.0168	0.1193	0.1324	0.1838
	Sundamage > 5	0.0576	0.0756	0.2230	0.5869	0.1134	<1e-04

Dispersion parameter Estimates	
Cluster frailty dispersion (BCC)	0.0126
Cluster frailty dispersion (SCC)	0.0016
Subject frailty dispersion (BCC)	0.7713
Subject frailty dispersion (SCC)	1.9495
Subject frailty covariance	0.3162
Subject frailty correlation	0.2557

(and largely quantitatively) unchanged. The subject-level frailty dispersion estimates are observed to decrease, indicating that the inclusion of this time-dependent covariate captures some of the subject-level heterogeneity.

Table 3.6 shows that during the course of the study, the average plasma selenium level in the treatment group was 48% higher than in the placebo group. Because treatment causes plasma selenium to rise considerably, the mean level of this covariate depends strongly on the treatment indicator. This may help to explain the strong similarity between the results summarized in Tables 3.8 and 3.9.

3.7.2 Effect of rhDNAse treatment on recurrent pulmonary exacerbations

In this section, we demonstrate the applicability of the proposed methodology to a bivariate process having a complex at-risk structure. The study consists of cystic fibrosis patients who experience numerous lengthy episodes of pulmonary exacerbation. The two recurrent event processes under consideration are the beginnings and endings of such episodes, or "alternating episode" data. At any point in time, a patient is considered to be at risk for exactly one of these event types, so the associated at-risk indicators alternate with each other. We discuss the findings of past analyses of these data in Section 3.7.2.1, discuss the methodology for our analysis in Section 3.7.2.2 and summarize the results in Section 3.7.2.3.

3.7.2.1 Study methods and findings

In a randomized double-blind study conducted by Genentech Inc. in 1992, a total of 968 cystic fibrosis patients in 51 institutions were followed for 24 weeks. Patients were eligible for the study if they were at least five years of age and had a confirmed diagnosis of cystic fibrosis. Randomization assigned 324 patients to placebo, 321 to be treated with 2.5mg of rhDNase once daily, and 343 to be treated with rhDNase twice daily (Fuchs et al., 1994). During exacerbation episodes, patients were treated with intravenous antibiotics and were not considered at risk for another episode until 7 days after the end of therapy. Observation periods for 17 patients terminated prematurely, that is, were censored prior to the end of the study.

Data on the placebo and once-daily treatment groups were made publicly available by Therneau and Grambsch (2000), and have been analyzed several times in

Table 3.10: Frequency distribution of the number of pulmonary exacerbation episodes observed.

Number of episodes:	0	1	2	3	4	5
Placebo	185	97	24	13	4	1
Treatment	217	65	30	6	3	0

Table 3.11: Average lengths of uncensored gap times, by episode and treatment group. Gap times considered are the at-risk time prior to each exacerbation starting, and the duration of the exacerbation itself.

	Exacerbation Start			Exacerbation End		
Episode	Plc	Trt	All	Plc	Trt	All
1	72.3	76.1	73.9	15.1	13.7	14.5
2	42.0	51.1	46.5	19.0	17.3	18.1
3	38.1	28.7	35.0	12.4	16.1	13.6
4	37.2	22.0	31.5	13.8	16.0	14.6
5	20.0	–	20.0	3.0	–	3.0
All	61.8	65.9	63.6	15.6	14.8	15.2

previous papers. For example, Lin et al. (2001) treated the arrival times of exacerbations as a recurrent event process, modeling the mean count as a known semiparametric transformation of a proportional means model; Cook and Lawless (2002) considered intensity models for the exacerbation lengths; and, Yan and Fine (2005) analyzed the number of exacerbations, the number of days in exacerbation, and the proportion of days in exacerbation using a novel temporal process regression approach.

Table 3.10 shows the frequency distribution of the number of exacerbations observed during the study. A total of 360 exacerbation episodes were observed, 205 and 155 in the placebo and treatment groups respectively. In the placebo group, 139 patients (42.9%) suffered at least one episode, versus 104 (32.4%) in the treatment group. Table 3.11 shows the average lengths of the uncensored intervals between exacerbations and the durations of the exacerbations themselves. For the first two exacerbations, the placebo group has shorter gap times between

exacerbations, and longer-lasting exacerbation episodes than the treatment group. This trend is reversed for the third and fourth episode; however, this may be an artifact of censoring, as relatively few patients experienced a third or fourth episode during the study period.

In addition to treatment information, the publicly available data set contains a single time-independent covariate. Forced expiratory volume (FEV) is a measure of a patient's lung capacity and often treated as a surrogate for the patient's overall health (Yan and Fine, 2005). The baseline FEV at study entry given in the data set ranges from 16 to 140.8 ml, with a mean of 61.1 ml and a standard deviation of 26.1. Patients who did not suffer any episodes had a mean FEV of 67.1 ml, whereas patients who suffered at least one had a mean FEV of 51.1 ml, indicating that FEV may have a strong effect on the risk of experiencing exacerbation episodes.

The original study of Fuchs et al. (1994) fits a Cox proportional hazards model to the time of first exacerbation only, with patient age as the sole covariate, and finds a statistically significant reduction of risk. Increase in FEV is treated as secondary endpoint. A more contemporary analysis by Therneau and Hamilton (1997) considers a proportional hazards model which treats the first and second episodes as distinct strata, and compares the results to an independent-increments model for exacerbation start times. Neither of the two approaches captures the dependence between multiple events for a single subject.

A later reanalysis of the data by Yan and Fine (2005) considers all exacerbation episodes, while allowing for time-dependent covariate effects and accounting for the discontinuous at-risk intervals. The "temporal process regression" methodology used in this paper fits GLMs to "snapshot" cross-sections of the data at each time, with responses given by either the number of exacerbations prior to that

time, or the cumulative or average time spent in exacerbation prior to that time. Their analyses detected significant time-dependent covariate effects: treatment in particular had a different effect on the first episode than on later ones. Because of the very different model structure, it is difficult to compare the results of their study to our own results presented in Section 3.7.2.3, however the finding of different covariate effects for each episode is very valuable.

3.7.2.2 Data analysis methodology

In their analysis, Yan and Fine (2005) expressed concern that previous approaches to episodic data using intensity models were unable to account properly for the unusual form of the at-risk function, and were unable to capture different effects of covariates on the alternating gap times. Our proposed methodology is capable of addressing both concerns.

We analyze episodic data as a bivariate recurrent event process, where the two events of interest are the beginning and end of an exacerbation. However, unlike the analysis for the data described in Section 3.7.1, patients are never simultaneously at risk for both types of event. Rather, the two at-risk processes alternate; hence, while exacerbated, patients are considered "at risk" for ending their exacerbation, but not for entering an exacerbation. Similarly, patients are only considered to be at risk for starting an exacerbation seven days after the end of their previous exacerbation episode.

For purposes of the data analysis, FEV was centered by its mean. Although the patients are clustered into 51 institutions, this information is not contained in the public data set; hence, we treat these data as if there were no clustering. Of

Table 3.12: Regression and dispersion parameter estimates for a basic bivariate frailty model fit to the Pulmonary Exacerbation data.

Regression parameter estimates						
	Exacerbation Start			Exacerbation End		
	Coeff	Std	Pval	Coeff	Std	Pval
Treatment	−0.3308	0.1396	0.0089	0.0645	0.1078	0.2750
FEV	−0.0203	0.0029	<1e-04	0.0105	0.0023	<1e-04

Dispersion parameter Estimates	
Subject frailty dispersion (Exacerbation Start)	1.4558
Subject frailty dispersion (Exacerbation End)	0.0652
Subject frailty covariance	-0.1302
Subject frailty correlation	-0.4229

the 645 patients in the placebo and once-daily treatment groups, 628 were followed until the end of the study; hence, it appears reasonable to assume that censoring is noninformative. Three models are considered. First, we include only treatment and FEV as covariates, assuming neither time- nor episode-dependence. Next, we add an indicator of whether the patient has suffered two or more previous episodes, as a means of accounting for effects specific to the first episodes. Lastly, we allow for episode-dependent effects of the two covariates, analogously to Yan and Fine (2005).

3.7.2.3 Data analysis results

Table 3.12 summarizes the results from fitting the data to the bivariate frailty model using Treatment and FEV as the only covariates. The panel labeled "Exacerbation Start" refers to the gap times in between exacerbations (i.e., "exacerbation free" periods); the panel labeled "Exacerbation End" refers to the lengths of exacerbations. The results indicate that patients in the treatment group and patients with higher FEV have a lower rate of starting a new exacerbation and,

Table 3.13: Regression and dispersion parameter estimates for a bivariate frailty model fit to the Pulmonary Exacerbation data, including the number of previous exacerbations as a covariate.

Regression parameter estimates						
	Exacerbation Start			Exacerbation End		
	Coeff	Std	Pval	Coeff	Std	Pval
Treatment	−0.3171	0.1367	0.0102	0.0693	0.1069	0.2586
FEV	−0.0197	0.0028	<1e-04	0.0105	0.0023	<1e-04
Past Ex. (2+)	0.2555	0.3080	0.2034	0.3877	0.1535	0.0058

Dispersion parameter Estimates	
Subject frailty dispersion (Exacerbation Start)	1.2779
Subject frailty dispersion (Exacerbation End)	0.0654
Subject frailty covariance	-0.1432
Subject frailty correlation	-0.4952

having started one, tend to end the current exacerbation more quickly. The effect of treatment on the exacerbation length (i.e., ending a current exacerbation) is statistically insignificant. This is consistent with Table 3.11, where it is observed that patients in the treatment group had shorter exacerbation episodes (and longer exacerbation-free periods) for the first two episodes, the pattern reversing itself for later episodes. The estimated dispersion parameters indicate a substantially higher level of heterogeneity in the rate at which exacerbations begin (i.e., in exacerbation-free periods) in comparison to the rate at which the current exacerbation ends. This may indicate that the covariates better explain the patient heterogeneity in the rate at which exacerbation end but not in exacerbation-free periods. The negative frailty covariance and correlation suggest that exacerbation and exacerbation-free periods are negatively correlated with each other even after accounting for covariate effects, a result that is not unexpected.

In order to allow for an episode-dependent effect, we next include an indicator for whether the patient has suffered two or more exacerbations during the course of the study. That is, the covariate takes value 0 before and during the first two

Table 3.14: Regression and dispersion parameter estimates for a bivariate frailty model fit to the Pulmonary Exacerbation data, with episode-dependent coefficients

Regression parameter estimates						
	Exacerbation Start			Exacerbation End		
	Coeff	Std	Pval	Coeff	Std	Pval
Treatment (Ep 1)	−0.4899	0.1433	0.0003	0.1313	0.1219	0.1406
Treatment (Ep 2)	−0.0303	0.2497	0.4517	−0.1292	0.1686	0.2217
Treatment (Ep 3-5)	0.0216	0.5244	0.4836	0.0088	0.2717	0.4871
FEV (Ep 1)	−0.0267	0.0031	<1e-04	0.0112	0.0030	<1e-04
FEV (Ep 2)	−0.0022	0.0070	0.3752	0.0087	0.0040	0.0137
FEV (Ep 3-5)	−0.0019	0.0146	0.4485	−0.0014	0.0060	0.4090

Dispersion parameter Estimates	
Subject frailty dispersion (Exacerbation Start)	1.1997
Subject frailty dispersion (Exacerbation End)	0.0649
Subject frailty covariance	-0.1150
Subject frailty correlation	-0.4123

exacerbation episodes, and takes value 1 thereafter. This choice of covariate was prompted by the evidence in Table 3.11 that the first and second episodes may have different characteristics from later episodes. Results of the fit are shown in Table 3.13. The estimated positive coefficient suggests that both later exacerbations and exacerbation-free periods tend to be shorter than earlier ones, perhaps suggesting an overall rise in the event frequency however, there is only a statistically significant effect on exacerbation periods. Also, adding in this covariate does not have a significant impact on any of the other parameter estimates summarized in Table 3.12.

Lastly, we consider a model in which the effects of Treatment and FEV are allowed to be different for each episode. Because of the rarity of third, fourth and fifth episodes, these events are grouped together into a single category. The results in Table 3.14 suggest that there is a need to consider episode-dependent covariate effects for these data. Treatment only has a statistically significant beneficial effect on the rate at which patients experience the first episode; it does not appear

to significantly impact later episodes or the length of exacerbation-free periods. Higher baseline FEV is observed to increase the length of the first exacerbation-free and subsequently decrease the length of the first exacerbation period, with these same effects persisting but less pronounced in the second pair of episodes. Covariate effects that are considerably stronger for earlier episodes are consistent with the time-dependent effects reported by Yan and Fine (2005). The dispersion parameters in Table 3.14, particularly the frailty covariance and correlation, are observed to have decreased slightly in comparison with previous fits, suggesting that the episode-dependent covariate helps to capture some of the subject-level heterogeneity in the event processes.

3.8 Discussion

The proposed methodology improves on existing methods for analyzing bivariate recurrent event data by not requiring a parametric specification of the frailty distribution. The resulting model is able to accommodate many features of real-world data, including clustering, stratification, unusual at-risk processes and time-dependent covariates. The accompanying `R` package `blupsurv` provides a useful new tool for statisticians for the analysis of both bivariate and univariate recurrent event data; we are not currently aware of another software package with the capability to fit such models.

Much like existing approaches, the proposed methodology requires significant computational resources. The computationally expensive numerical integrations required for likelihood-based estimation under a specified frailty distribution are avoided, but only at the cost of having to estimate a potentially large number of

nuisance parameters. The use discretization offsets these computational costs and, at the expense of some bias, leads to a relatively fast method of estimation for all model parameters and tractable computation of standard errors. Nevertheless, the method's complexity remains such that model selection with large datasets may require many hours on the average personal computer or workstation.

Our simulation results suggest that the methodology leads to consistent estimates of regression and frailty dispersion parameters, including the frailty covariance. The general similarity of the model to that considered in Ma and Jørgensen (2007) suggests that the desired large-sample properties may hold under a suitable asymptotic framework. However, a proof of this fact has proved to be elusive and no asymptotic justification for the proposed approach is currently available.

APPENDIX

Appendix 3.A Additional Simulation Results

The following tables contain simulation results that accompany or extend those given in Sec. 3.6. Tables 3.A–3.A contain biases, means and variances for the three simulation settings considered.

Table 3.A contains simulation results using the marginal estimators in eq. (3.5.2), under setting (I). The results indicate that the marginal estimators perform nearly as well as the Pearson estimators, although very fine discretization sometimes causes the correlation coefficient to be overestimated.

Table 3.A demonstrates the performance of the method with time-dependent covariates. Simulated data are generated using a single time-dependent covariate that changes at random intervals, and the model is fit using the approximation presented in the paper. Results indicate that the model performs well with time-dependent covariates: biases are only slightly higher than with fixed covariates, and the error can be accounted for inexactness in the discretization.

Lastly, Table 3.21 shows the effect of the degree-of-freedom adjustment suggested by Ma (1999). The adjustment tends to lead to overestimation of the subject-level dispersion parameters, especially at fine levels of discretization, however estimates of the cluster-level dispersion parameters are much improved.

Table 3.15: Full parameter estimation results from 1000 simulations under setting (I), matching table 3.3.

	m	J_i	K	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\nu}_1^2$	$\hat{\nu}_2^2$	$\hat{\theta}$	$\hat{\rho}$
Bias	10	5	.1	−0.076	−0.077	−0.130	−0.135	−0.146	−0.148	−0.073	0.001
			.25	−0.033	−0.024	−0.076	−0.084	−0.089	−0.099	−0.047	0.008
			.5	−0.022	−0.007	−0.079	−0.064	−0.069	−0.050	−0.032	0.029
			1	−0.000	−0.010	−0.072	−0.061	−0.062	−0.065	−0.028	0.027
		25	.1	−0.024	−0.024	−0.056	−0.057	−0.065	−0.066	−0.022	0.046
			.25	−0.005	−0.007	−0.043	−0.044	−0.037	−0.039	−0.008	0.050
			.5	−0.005	0.001	−0.054	−0.056	−0.043	−0.038	−0.010	0.051
			1	−0.001	0.001	−0.051	−0.063	−0.056	−0.057	−0.017	0.044
	25	5	.1	−0.046	−0.046	−0.066	−0.069	−0.102	−0.111	−0.046	0.002
			.25	−0.023	−0.023	−0.039	−0.045	−0.071	−0.070	−0.031	0.005
			.5	−0.014	−0.008	−0.051	−0.050	−0.062	−0.057	−0.022	0.024
			1	−0.001	−0.010	−0.042	−0.044	−0.057	−0.058	−0.024	0.012
		25	.1	−0.008	−0.013	−0.024	−0.024	−0.047	−0.051	−0.015	0.019
			.25	−0.001	−0.006	−0.021	−0.022	−0.040	−0.045	−0.010	0.023
			.5	−0.002	0.000	−0.030	−0.025	−0.047	−0.051	−0.014	0.021
			1	−0.003	−0.000	−0.032	−0.033	−0.061	−0.061	−0.020	0.017
Var	10	5	.1	0.059	0.061	0.016	0.012	0.006	0.007	0.003	0.053
			.25	0.069	0.063	0.084	0.034	0.043	0.013	0.006	0.054
			.5	0.070	0.074	0.032	0.050	0.019	0.043	0.008	0.058
			1	0.078	0.072	0.028	0.046	0.026	0.020	0.008	0.050
		25	.1	0.011	0.011	0.027	0.023	0.007	0.007	0.002	0.014
			.25	0.011	0.011	0.021	0.021	0.009	0.009	0.002	0.014
			.5	0.011	0.010	0.018	0.019	0.007	0.007	0.002	0.014
			1	0.013	0.012	0.019	0.016	0.005	0.005	0.002	0.014
	25	5	.1	0.025	0.023	0.016	0.010	0.006	0.004	0.002	0.020
			.25	0.028	0.025	0.020	0.036	0.009	0.009	0.003	0.019
			.5	0.027	0.026	0.013	0.015	0.008	0.009	0.003	0.021
			1	0.027	0.027	0.016	0.016	0.009	0.009	0.003	0.019
		25	.1	0.004	0.004	0.012	0.013	0.003	0.003	0.000	0.005
			.25	0.004	0.005	0.012	0.012	0.003	0.003	0.001	0.005
			.5	0.004	0.005	0.011	0.011	0.003	0.002	0.000	0.005
			1	0.004	0.004	0.012	0.009	0.002	0.002	0.000	0.005
MSE	10	5	.1	0.065	0.067	0.033	0.030	0.028	0.029	0.008	0.053
			.25	0.070	0.064	0.090	0.042	0.050	0.023	0.009	0.054
			.5	0.070	0.074	0.038	0.055	0.024	0.046	0.009	0.059
			1	0.078	0.072	0.033	0.049	0.029	0.024	0.009	0.051
		25	.1	0.011	0.012	0.030	0.026	0.012	0.012	0.002	0.016
			.25	0.011	0.011	0.023	0.023	0.011	0.011	0.002	0.016
			.5	0.011	0.010	0.021	0.022	0.009	0.009	0.002	0.017
			1	0.013	0.012	0.022	0.020	0.009	0.009	0.002	0.016
	25	5	.1	0.027	0.025	0.020	0.015	0.016	0.017	0.004	0.020
			.25	0.028	0.026	0.021	0.038	0.014	0.014	0.004	0.019
			.5	0.027	0.026	0.016	0.017	0.012	0.013	0.004	0.022
			1	0.027	0.027	0.018	0.018	0.012	0.012	0.004	0.019
		25	.1	0.004	0.004	0.012	0.014	0.005	0.005	0.001	0.005
			.25	0.004	0.005	0.013	0.013	0.005	0.005	0.001	0.006
			.5	0.004	0.005	0.012	0.012	0.005	0.005	0.001	0.005
			1	0.004	0.004	0.013	0.010	0.006	0.005	0.001	0.006

Table 3.16: Full parameter estimation results from 1000 simulations under setting (II), matching table 3.4.

	m	J_i	ϕ	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\sigma}^2_{(0)}$	$\hat{\sigma}^2_{(1)}$	$\hat{\nu}^2_{(0)}$	$\hat{\nu}^2_{(1)}$	$\hat{\theta}$	$\hat{\rho}$
Bias	10	5	.1	−0.069	−0.140	−0.127	−0.136	−0.146	−0.149	−0.072	0.000
			.25	−0.034	−0.065	−0.089	−0.087	−0.081	−0.089	−0.039	0.036
			.5	−0.011	−0.033	−0.080	−0.074	−0.062	−0.060	−0.029	0.038
			1	0.007	−0.004	−0.074	−0.070	−0.068	−0.058	−0.035	0.018
		25	.1	−0.029	−0.045	−0.059	−0.067	−0.070	−0.075	−0.024	0.050
			.25	−0.002	−0.018	−0.048	−0.051	−0.038	−0.049	−0.012	0.049
			.5	−0.010	−0.005	−0.056	−0.055	−0.045	−0.046	−0.011	0.050
			1	−0.004	−0.007	−0.052	−0.070	−0.062	−0.062	−0.020	0.043
	25	5	.1	−0.052	−0.089	−0.073	−0.080	−0.102	−0.097	−0.042	0.016
			.25	−0.016	−0.032	−0.053	−0.045	−0.069	−0.074	−0.029	0.014
			.5	−0.013	−0.022	−0.043	−0.049	−0.056	−0.058	−0.022	0.020
			1	−0.005	−0.002	−0.045	−0.049	−0.056	−0.062	−0.025	0.013
		25	.1	−0.010	−0.022	−0.017	−0.031	−0.043	−0.057	−0.015	0.019
			.25	−0.003	−0.004	−0.025	−0.025	−0.043	−0.046	−0.012	0.023
			.5	−0.004	−0.007	−0.024	−0.025	−0.049	−0.050	−0.014	0.018
			1	−0.002	−0.003	−0.039	−0.037	−0.063	−0.060	−0.020	0.020
Var	10	5	.1	0.062	0.083	0.017	0.011	0.006	0.006	0.003	0.056
			.25	0.066	0.084	0.031	0.038	0.109	0.022	0.006	0.054
			.5	0.072	0.082	0.037	0.034	0.024	0.025	0.008	0.058
			1	0.070	0.092	0.028	0.029	0.020	0.030	0.008	0.055
		25	.1	0.011	0.014	0.023	0.023	0.005	0.005	0.001	0.014
			.25	0.012	0.015	0.019	0.023	0.008	0.007	0.002	0.014
			.5	0.012	0.015	0.016	0.016	0.007	0.006	0.002	0.013
			1	0.012	0.015	0.020	0.014	0.005	0.005	0.002	0.014
	25	5	.1	0.023	0.028	0.011	0.010	0.005	0.006	0.002	0.021
			.25	0.026	0.032	0.017	0.017	0.009	0.008	0.003	0.019
			.5	0.026	0.031	0.021	0.014	0.011	0.009	0.003	0.019
			1	0.026	0.036	0.016	0.014	0.010	0.008	0.003	0.020
		25	.1	0.004	0.005	0.014	0.011	0.003	0.002	0.000	0.005
			.25	0.004	0.006	0.012	0.013	0.002	0.002	0.000	0.005
			.5	0.004	0.006	0.012	0.012	0.003	0.002	0.000	0.005
			1	0.004	0.006	0.008	0.008	0.002	0.002	0.000	0.005
MSE	10	5	.1	0.067	0.103	0.033	0.030	0.028	0.029	0.008	0.056
			.25	0.067	0.088	0.039	0.045	0.116	0.030	0.008	0.055
			.5	0.072	0.083	0.044	0.040	0.028	0.029	0.009	0.059
			1	0.070	0.092	0.034	0.034	0.024	0.033	0.009	0.055
		25	.1	0.012	0.016	0.026	0.028	0.010	0.011	0.002	0.017
			.25	0.012	0.016	0.022	0.025	0.009	0.010	0.002	0.016
			.5	0.012	0.015	0.019	0.019	0.009	0.008	0.002	0.015
			1	0.012	0.015	0.023	0.019	0.009	0.009	0.002	0.015
	25	5	.1	0.026	0.036	0.017	0.017	0.016	0.015	0.004	0.021
			.25	0.026	0.033	0.020	0.019	0.014	0.014	0.003	0.019
			.5	0.026	0.031	0.023	0.017	0.014	0.012	0.003	0.019
			1	0.026	0.036	0.018	0.016	0.013	0.012	0.003	0.020
		25	.1	0.005	0.006	0.015	0.012	0.005	0.005	0.001	0.006
			.25	0.004	0.006	0.012	0.014	0.004	0.005	0.001	0.005
			.5	0.004	0.006	0.012	0.013	0.005	0.004	0.001	0.005
			1	0.004	0.006	0.009	0.010	0.006	0.005	0.001	0.006

Table 3.17: Full parameter estimation results from 1000 simulations under setting (III), matching table 3.5.

	m	J_i	ϕ	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\sigma}^2_{(0)}$	$\hat{\sigma}^2_{(1)}$	$\hat{\nu}^2_{(0)}$	$\hat{\nu}^2_{(1)}$	$\hat{\theta}$	$\hat{\rho}$
Bias	10	5	.1	−0.082	−0.125	−0.124	−0.287	−0.151	−0.325	−0.065	0.036
			.25	−0.014	−0.068	−0.080	−0.164	−0.084	−0.147	−0.025	0.057
			.5	−0.027	−0.035	−0.066	−0.138	−0.069	−0.136	−0.024	0.046
			1	0.011	−0.028	−0.060	−0.153	−0.065	−0.129	−0.015	0.057
		25	.1	−0.035	−0.032	−0.046	−0.068	−0.070	−0.029	−0.001	0.062
			.25	−0.014	−0.007	−0.041	−0.172	−0.043	−0.042	0.007	0.074
			.5	−0.010	−0.007	−0.046	−0.178	−0.043	−0.040	0.004	0.067
			1	−0.006	−0.006	−0.052	−0.180	−0.063	−0.076	−0.004	0.061
	25	5	.1	−0.044	−0.066	−0.065	−0.129	−0.108	−0.194	−0.032	0.034
			.25	−0.028	−0.034	−0.028	−0.064	−0.065	−0.121	−0.012	0.033
			.5	−0.015	−0.018	−0.033	−0.057	−0.062	−0.095	−0.014	0.026
			1	−0.018	−0.019	−0.035	−0.076	−0.060	−0.111	−0.012	0.033
		25	.1	−0.012	−0.005	−0.015	−0.056	−0.042	0.035	0.012	0.041
			.25	−0.005	0.009	−0.009	−0.086	−0.045	0.012	0.007	0.037
			.5	−0.007	0.006	−0.016	−0.084	−0.049	−0.022	0.002	0.035
			1	−0.004	0.004	−0.025	−0.082	−0.063	−0.083	−0.006	0.031
Var	10	5	.1	0.063	0.081	0.013	0.033	0.006	0.026	0.004	0.049
			.25	0.074	0.089	0.033	0.203	0.030	0.334	0.020	0.055
			.5	0.072	0.100	0.035	0.215	0.022	0.108	0.012	0.049
			1	0.069	0.098	0.035	0.177	0.019	0.114	0.012	0.049
		25	.1	0.011	0.017	0.023	0.137	0.007	0.072	0.003	0.013
			.25	0.011	0.016	0.023	0.059	0.009	0.047	0.004	0.013
			.5	0.011	0.017	0.018	0.039	0.007	0.051	0.004	0.013
			1	0.011	0.017	0.021	0.037	0.006	0.043	0.003	0.011
	25	5	.1	0.026	0.035	0.013	0.078	0.007	0.037	0.004	0.016
			.25	0.026	0.034	0.023	0.109	0.012	0.050	0.006	0.017
			.5	0.028	0.037	0.018	0.108	0.009	0.070	0.005	0.016
			1	0.027	0.039	0.015	0.081	0.011	0.055	0.005	0.017
		25	.1	0.004	0.007	0.014	0.035	0.004	0.038	0.001	0.004
			.25	0.005	0.006	0.013	0.029	0.004	0.028	0.001	0.004
			.5	0.004	0.007	0.012	0.030	0.002	0.024	0.001	0.004
			1	0.004	0.006	0.010	0.031	0.002	0.014	0.001	0.004
MSE	10	5	.1	0.070	0.097	0.028	0.116	0.029	0.132	0.009	0.051
			.25	0.074	0.094	0.039	0.230	0.037	0.355	0.020	0.058
			.5	0.072	0.101	0.039	0.234	0.027	0.127	0.013	0.051
			1	0.069	0.099	0.038	0.200	0.024	0.130	0.012	0.052
		25	.1	0.013	0.018	0.025	0.142	0.012	0.073	0.003	0.017
			.25	0.011	0.016	0.024	0.089	0.011	0.049	0.004	0.019
			.5	0.011	0.017	0.020	0.070	0.008	0.053	0.004	0.017
			1	0.011	0.017	0.024	0.069	0.010	0.049	0.003	0.015
	25	5	.1	0.028	0.039	0.017	0.095	0.018	0.075	0.005	0.018
			.25	0.027	0.035	0.024	0.113	0.016	0.064	0.006	0.018
			.5	0.029	0.037	0.019	0.111	0.013	0.079	0.006	0.017
			1	0.027	0.039	0.016	0.086	0.014	0.067	0.005	0.018
		25	.1	0.004	0.007	0.014	0.038	0.006	0.039	0.002	0.006
			.25	0.005	0.006	0.013	0.036	0.006	0.028	0.001	0.005
			.5	0.004	0.007	0.012	0.038	0.005	0.025	0.001	0.005
			1	0.004	0.006	0.011	0.038	0.006	0.021	0.001	0.005

Table 3.18: Parameter estimation results from 1000 simulations for two recurrent event processes, with dispersion parameters estimated using the marginal estimators in (3.5.2)

	m	J_i	ϕ	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\sigma}^2_{(0)}$	$\hat{\sigma}^2_{(1)}$	$\hat{\nu}^2_{(0)}$	$\hat{\nu}^2_{(1)}$	$\hat{\theta}$	$\hat{\rho}$
Bias	10	5	.1	−0.084	−0.066	−0.140	−0.134	−0.153	−0.151	−0.086	−0.057
			.25	−0.045	−0.039	−0.124	−0.125	−0.130	−0.131	−0.071	−0.026
			.5	−0.027	−0.031	−0.107	−0.114	−0.121	−0.113	−0.063	−0.024
			1	−0.001	−0.002	−0.096	−0.103	−0.107	−0.112	−0.055	−0.007
		25	.1	−0.038	−0.039	−0.101	−0.105	−0.114	−0.110	−0.042	0.050
			.25	−0.026	−0.028	−0.089	−0.091	−0.100	−0.098	−0.035	0.049
			.5	−0.015	−0.014	−0.084	−0.081	−0.092	−0.092	−0.028	0.055
			1	−0.017	−0.011	−0.083	−0.077	−0.082	−0.084	−0.019	0.073
	25	5	.1	−0.061	−0.058	−0.106	−0.111	−0.128	−0.128	−0.056	0.014
			.25	−0.029	−0.043	−0.091	−0.088	−0.114	−0.113	−0.050	0.008
			.5	−0.026	−0.031	−0.083	−0.082	−0.103	−0.107	−0.043	0.014
			1	−0.010	−0.023	−0.072	−0.075	−0.098	−0.097	−0.038	0.016
		25	.1	−0.026	−0.029	−0.079	−0.079	−0.101	−0.102	−0.030	0.051
			.25	−0.015	−0.018	−0.070	−0.072	−0.090	−0.091	−0.023	0.056
			.5	−0.018	−0.016	−0.067	−0.064	−0.086	−0.083	−0.015	0.068
			1	−0.013	−0.015	−0.059	−0.063	−0.077	−0.077	−0.012	0.067
Var	10	5	.1	0.058	0.064	0.007	0.008	0.006	0.006	0.003	0.072
			.25	0.061	0.069	0.009	0.009	0.010	0.009	0.005	0.083
			.5	0.070	0.076	0.012	0.010	0.010	0.011	0.007	0.085
			1	0.069	0.074	0.012	0.013	0.011	0.010	0.006	0.073
		25	.1	0.012	0.010	0.005	0.005	0.002	0.002	0.003	0.043
			.25	0.011	0.012	0.006	0.006	0.002	0.003	0.003	0.038
			.5	0.011	0.012	0.006	0.006	0.003	0.002	0.003	0.036
			1	0.011	0.010	0.005	0.006	0.002	0.002	0.003	0.033
	25	5	.1	0.025	0.022	0.004	0.004	0.004	0.004	0.002	0.038
			.25	0.025	0.023	0.005	0.006	0.004	0.005	0.003	0.042
			.5	0.025	0.024	0.005	0.006	0.005	0.005	0.004	0.040
			1	0.025	0.026	0.006	0.006	0.005	0.005	0.004	0.035
		25	.1	0.004	0.004	0.002	0.002	0.001	0.001	0.001	0.017
			.25	0.004	0.004	0.002	0.002	0.001	0.001	0.002	0.018
			.5	0.004	0.004	0.002	0.002	0.001	0.001	0.001	0.015
			1	0.004	0.004	0.003	0.003	0.001	0.001	0.002	0.016
MSE	10	5	.1	0.065	0.069	0.027	0.026	0.030	0.029	0.010	0.075
			.25	0.063	0.070	0.025	0.025	0.027	0.026	0.011	0.084
			.5	0.071	0.076	0.023	0.024	0.025	0.024	0.011	0.086
			1	0.069	0.074	0.021	0.023	0.022	0.023	0.010	0.073
		25	.1	0.014	0.012	0.015	0.016	0.015	0.014	0.005	0.045
			.25	0.012	0.013	0.014	0.014	0.012	0.012	0.005	0.040
			.5	0.011	0.012	0.013	0.013	0.011	0.011	0.004	0.039
			1	0.011	0.011	0.012	0.012	0.009	0.010	0.004	0.039
	25	5	.1	0.028	0.026	0.016	0.017	0.020	0.020	0.006	0.038
			.25	0.026	0.025	0.014	0.014	0.017	0.018	0.006	0.042
			.5	0.026	0.025	0.012	0.013	0.016	0.017	0.006	0.040
			1	0.025	0.026	0.011	0.012	0.014	0.014	0.005	0.035
		25	.1	0.005	0.005	0.008	0.008	0.011	0.011	0.002	0.019
			.25	0.004	0.005	0.007	0.007	0.009	0.009	0.002	0.021
			.5	0.004	0.004	0.007	0.006	0.008	0.008	0.002	0.020
			1	0.004	0.005	0.006	0.007	0.007	0.007	0.002	0.020

Table 3.19: Parameter estimation results from 1000 simulations for two recurrent event processes with time-dependent covariates changing at Weibull intervals..

	m	J_i	ϕ	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\sigma}^2_{(0)}$	$\hat{\sigma}^2_{(1)}$	$\hat{\nu}^2_{(0)}$	$\hat{\nu}^2_{(1)}$	$\hat{\theta}$	$\hat{\rho}$
Bias	10	5	.1	−0.119	−0.113	−0.153	−0.147	−0.173	−0.173	−0.085	−0.012
			.25	−0.064	−0.059	−0.113	−0.111	−0.113	−0.124	−0.058	0.016
			.5	−0.023	−0.023	−0.094	−0.096	−0.083	−0.092	−0.041	0.029
			1	0.017	0.006	−0.088	−0.087	−0.063	−0.065	−0.031	0.032
		25	.1	−0.051	−0.044	−0.105	−0.098	−0.129	−0.132	−0.052	0.032
			.25	−0.030	−0.021	−0.073	−0.073	−0.090	−0.086	−0.031	0.042
			.5	−0.010	−0.010	−0.071	−0.069	−0.065	−0.065	−0.019	0.057
			1	−0.003	−0.003	−0.062	−0.065	−0.072	−0.076	−0.025	0.041
	25	5	.1	−0.071	−0.077	−0.098	−0.099	−0.130	−0.133	−0.059	−0.001
			.25	−0.035	−0.032	−0.070	−0.067	−0.093	−0.089	−0.040	0.008
			.5	−0.012	−0.012	−0.054	−0.048	−0.075	−0.072	−0.032	0.007
			1	0.003	0.010	−0.051	−0.052	−0.065	−0.068	−0.026	0.018
		25	.1	−0.028	−0.027	−0.062	−0.057	−0.101	−0.097	−0.034	0.023
			.25	−0.014	−0.012	−0.044	−0.034	−0.068	−0.070	−0.021	0.024
			.5	−0.007	−0.007	−0.042	−0.034	−0.063	−0.061	−0.016	0.032
			1	0.000	−0.005	−0.040	−0.038	−0.068	−0.066	−0.019	0.027
Var	10	5	.1	0.034	0.033	0.006	0.007	0.003	0.003	0.001	0.048
			.25	0.037	0.035	0.018	0.019	0.011	0.009	0.003	0.049
			.5	0.035	0.035	0.021	0.031	0.018	0.011	0.006	0.053
			1	0.031	0.037	0.025	0.020	0.021	0.020	0.008	0.055
		25	.1	0.006	0.007	0.008	0.009	0.002	0.002	0.001	0.014
			.25	0.006	0.005	0.015	0.016	0.004	0.005	0.001	0.013
			.5	0.006	0.006	0.014	0.015	0.007	0.006	0.001	0.013
			1	0.007	0.006	0.018	0.016	0.005	0.005	0.001	0.013
	25	5	.1	0.012	0.012	0.007	0.007	0.003	0.003	0.001	0.020
			.25	0.014	0.012	0.014	0.013	0.006	0.007	0.002	0.020
			.5	0.014	0.013	0.012	0.015	0.007	0.007	0.003	0.019
			1	0.013	0.013	0.013	0.013	0.008	0.007	0.003	0.019
		25	.1	0.002	0.002	0.006	0.008	0.001	0.001	0.000	0.005
			.25	0.002	0.002	0.008	0.010	0.002	0.002	0.000	0.006
			.5	0.002	0.002	0.008	0.010	0.002	0.002	0.000	0.005
			1	0.002	0.002	0.009	0.009	0.001	0.002	0.000	0.005
MSE	10	5	.1	0.048	0.046	0.030	0.029	0.033	0.033	0.009	0.048
			.25	0.041	0.039	0.031	0.031	0.024	0.024	0.007	0.049
			.5	0.035	0.035	0.030	0.041	0.025	0.019	0.008	0.054
			1	0.032	0.037	0.033	0.028	0.025	0.025	0.009	0.056
		25	.1	0.009	0.008	0.019	0.019	0.019	0.020	0.003	0.015
			.25	0.007	0.006	0.020	0.021	0.012	0.012	0.002	0.015
			.5	0.006	0.006	0.019	0.020	0.011	0.010	0.002	0.017
			1	0.007	0.006	0.022	0.020	0.010	0.011	0.002	0.015
		5	.1	0.017	0.018	0.017	0.017	0.020	0.021	0.005	0.020
			.25	0.015	0.013	0.019	0.017	0.015	0.015	0.004	0.020
			.5	0.014	0.013	0.015	0.017	0.012	0.013	0.004	0.019
			1	0.013	0.013	0.016	0.016	0.012	0.012	0.003	0.019
		25	.1	0.003	0.003	0.009	0.011	0.011	0.011	0.001	0.006
			.25	0.002	0.002	0.010	0.011	0.006	0.007	0.001	0.006
			.5	0.002	0.002	0.010	0.011	0.006	0.006	0.001	0.006
			1	0.002	0.002	0.011	0.010	0.006	0.006	0.001	0.006

Table 3.20: Standard error estimation results of 1000 simulations for two recurrent event processes with time-dependent covariates

			$\hat{\beta}^{(0)}$			$\hat{\beta}^{(1)}$		
m	J_i	ϕ	CSE	ESE	95%CP	CSE	ESE	95%CP
10	5	.1	0.169	0.185	0.868	0.168	0.182	0.880
		.25	0.169	0.192	0.899	0.169	0.188	0.905
		.5	0.170	0.187	0.926	0.170	0.187	0.922
		1	0.168	0.178	0.932	0.168	0.193	0.909
	25	.1	0.071	0.079	0.861	0.071	0.083	0.856
		.25	0.073	0.079	0.902	0.073	0.075	0.921
		.5	0.073	0.081	0.931	0.073	0.079	0.926
		1	0.072	0.084	0.903	0.072	0.081	0.915
25	5	.1	0.105	0.110	0.885	0.105	0.112	0.865
		.25	0.106	0.118	0.909	0.106	0.111	0.918
		.5	0.105	0.118	0.915	0.105	0.115	0.925
		1	0.105	0.115	0.923	0.105	0.116	0.918
	25	.1	0.045	0.049	0.883	0.045	0.049	0.891
		.25	0.046	0.050	0.915	0.046	0.048	0.938
		.5	0.046	0.050	0.929	0.046	0.051	0.909
		1	0.046	0.049	0.924	0.046	0.051	0.926

Table 3.21: Parameter estimation results from 1000 simulations for two recurrent event processes with the small-sample bias correction proposed in Ma (1999) and presented in Section 3.5.1

	m	J_i	ϕ	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\sigma}^2_{(0)}$	$\hat{\sigma}^2_{(1)}$	$\hat{\nu}^2_{(0)}$	$\hat{\nu}^2_{(1)}$	$\hat{\theta}$	$\hat{\rho}$
Bias	10	5	.1	−0.047	−0.058	−0.088	−0.100	0.001	−0.001	−0.049	−0.057
			.25	0.009	0.011	−0.015	−0.029	0.106	0.098	−0.016	−0.051
			.5	0.017	0.019	−0.012	−0.016	0.143	0.138	−0.003	−0.044
			1	0.054	0.055	−0.029	−0.023	0.144	0.140	0.001	−0.034
		25	.1	0.020	0.019	0.015	0.030	0.168	0.161	0.015	−0.028
			.25	0.028	0.038	−0.003	0.006	0.189	0.200	0.026	−0.020
			.5	0.038	0.030	−0.016	−0.017	0.181	0.182	0.023	−0.017
			1	0.040	0.039	−0.011	−0.019	0.146	0.138	0.008	−0.027
	25	5	.1	−0.034	−0.034	−0.053	−0.057	−0.033	−0.030	−0.035	−0.024
			.25	−0.003	0.007	−0.024	−0.034	0.011	0.014	−0.015	−0.016
			.5	0.010	0.017	−0.023	−0.027	0.022	0.026	−0.017	−0.024
			1	0.024	0.001	−0.029	−0.037	0.015	0.019	−0.018	−0.023
		25	.1	0.013	0.014	0.015	0.020	0.051	0.053	0.004	−0.011
			.25	0.021	0.022	0.007	0.003	0.047	0.043	−0.000	−0.014
			.5	0.019	0.020	−0.010	−0.005	0.032	0.034	−0.002	−0.009
			1	0.015	0.021	−0.017	−0.014	0.015	0.017	−0.009	−0.013
Var	10	5	.1	0.064	0.066	0.021	0.021	0.012	0.011	0.005	0.026
			.25	0.072	0.066	0.079	0.056	0.050	0.045	0.011	0.026
			.5	0.072	0.076	0.055	0.089	0.055	0.061	0.017	0.030
			1	0.083	0.075	0.040	0.045	0.064	0.046	0.014	0.030
		25	.1	0.013	0.011	0.061	0.074	0.022	0.021	0.003	0.007
			.25	0.012	0.013	0.034	0.038	0.018	0.020	0.004	0.007
			.5	0.014	0.013	0.025	0.024	0.016	0.015	0.003	0.008
			1	0.013	0.012	0.029	0.025	0.013	0.011	0.003	0.007
	25	5	.1	0.025	0.026	0.017	0.015	0.007	0.008	0.002	0.014
			.25	0.029	0.025	0.025	0.017	0.011	0.012	0.003	0.015
			.5	0.025	0.029	0.024	0.024	0.013	0.013	0.004	0.016
			1	0.028	0.027	0.018	0.015	0.011	0.011	0.004	0.015
		25	.1	0.004	0.005	0.024	0.022	0.005	0.006	0.001	0.004
			.25	0.004	0.004	0.017	0.016	0.005	0.004	0.001	0.004
			.5	0.004	0.004	0.014	0.014	0.003	0.004	0.001	0.004
			1	0.005	0.004	0.011	0.010	0.002	0.003	0.001	0.004
MSE	10	5	.1	0.066	0.069	0.029	0.031	0.012	0.011	0.007	0.030
			.25	0.072	0.066	0.079	0.056	0.061	0.055	0.012	0.028
			.5	0.072	0.077	0.055	0.089	0.076	0.080	0.017	0.032
			1	0.086	0.078	0.041	0.045	0.085	0.066	0.014	0.031
		25	.1	0.013	0.012	0.061	0.075	0.050	0.048	0.004	0.008
			.25	0.013	0.015	0.034	0.038	0.054	0.060	0.004	0.008
			.5	0.015	0.014	0.025	0.024	0.049	0.048	0.004	0.008
			1	0.014	0.014	0.029	0.025	0.034	0.030	0.003	0.008
	25	5	.1	0.026	0.027	0.020	0.018	0.008	0.009	0.004	0.014
			.25	0.029	0.025	0.026	0.019	0.011	0.012	0.004	0.015
			.5	0.025	0.029	0.024	0.024	0.014	0.014	0.004	0.016
			1	0.028	0.027	0.019	0.016	0.011	0.012	0.004	0.015
		25	.1	0.005	0.005	0.024	0.022	0.008	0.009	0.001	0.004
			.25	0.005	0.005	0.017	0.016	0.007	0.006	0.001	0.004
			.5	0.005	0.005	0.014	0.014	0.004	0.005	0.001	0.004
			1	0.005	0.005	0.011	0.010	0.002	0.003	0.001	0.004

Table 3.22: Standard error estimation results from 1000 simulations for two recurrent event processes with the small-sample bias correction proposed in Ma (1999) and presented in 3.5.1

			$\hat{\beta}^{(0)}$			$\hat{\beta}^{(1)}$		
m	J_i	ϕ	CSE	ESE	95%CP	CSE	ESE	95%CP
10	5	.1	0.248	0.254	0.942	0.246	0.257	0.930
		.25	0.271	0.269	0.957	0.270	0.258	0.956
		.5	0.279	0.268	0.966	0.278	0.277	0.951
		1	0.279	0.289	0.936	0.278	0.273	0.949
	25	.1	0.120	0.114	0.955	0.119	0.109	0.962
		.25	0.122	0.113	0.966	0.123	0.117	0.951
		.5	0.121	0.118	0.956	0.121	0.115	0.957
		1	0.117	0.114	0.934	0.116	0.112	0.948
25	5	.1	0.150	0.159	0.931	0.151	0.162	0.927
		.25	0.158	0.171	0.922	0.159	0.159	0.948
		.5	0.159	0.160	0.949	0.159	0.170	0.931
		1	0.158	0.168	0.939	0.158	0.167	0.940
	25	.1	0.069	0.069	0.949	0.069	0.071	0.946
		.25	0.069	0.070	0.926	0.068	0.069	0.930
		.5	0.068	0.070	0.934	0.068	0.069	0.933
		1	0.066	0.070	0.937	0.066	0.067	0.932

BIBLIOGRAPHY

Aalen, O. and Husebye, E. (1991). Statistical analysis of repeated events forming renewal processes. *Statistics in Medicine* **10,** 1227–1240.

Abu-Libdeh, H., Turnbull, B. W., and Clark, L. C. (1990). Analysis of multi-type recurrent events in longitudinal studies; application to a skin cancer prevention trial. *Biometrics* **46,** 1017–1034.

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10,** 1100–1120.

Barker, P. and Henderson, R. (2005). Small sample bias in the gamma frailty model for univariate survival. *Lifetime Data Analysis* **11,**.

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

Breslow, N. E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82,** 81–91.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage i bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In M. Pavone-Macaluso, P. S. and Fedsmyn, F., editors, *Bladder Tumors and Other Topics in Urological Oncology*, pages 363–370. Plenum, New York.

Cai, J. and Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis* **10,** 121–138.

Chang, S.-H. and Wang, M.-C. (1999). Conditional regression analysis for recurrence time data. *Journal of the American Statistical Association* **94,** 1221–1230.

Clark, L. C., Combs, G. F. J., Turnbull, B. W., Slate, E. H., Chalker, D. K., Chow, J., Davis, L. S., Glover, R. A., Graham, G. F., Gross, E. G., Krongrad, A., Lesher, J. L. J., Park, H. K., Sanders, B. B. J., Smith, C. L., and Taylor, J. R. (1996). Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin. a randomized controlled trial. nutritional prevention of cancer study group. *JAMA* **276,** 1957–1963.

Cook, R. and Lawless, J. (2002). Analysis of repeated events. *Statistical Methods for Medical Research* **11,** 141–166.

Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Event Data*. Springer-Verlag, New York.

Cook, R. J., Ng, E., Mukherjee, J., and Vaughn, D. (1999). Two-state mixed renewal processes for chronic disease. *Statistics in Medicine* **18,** 175–188.

Cox, D. R. (1972a). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34,** 187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.

Cox, D. R. (1972b). The statistical analysis of dependencies in point processes. In Lewis, P., editor, *Stochastic Point Processes*, pages 55–66. Wiley, New York.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Duchateau, L., Janssen, P., Kezic, I., and Fortpied, C. (2003). Evolution of recurrent asthma event rate over time in frailty models. *J. Roy. Statist. Soc. Ser. C* **52,** 355–363.

Duffield-Lillico, A. J., Reid, M. E., Turnbull, B. W., Combs, G. F. J., Slate, E. H., Fischbach, L. A., Marshall, J. R., and Clark, L. C. (2002). Baseline characteristics and the effect of selenium supplementation on cancer incidence in a randomized clinical trial: a summary report of the nutritional prevention of cancer trial. *Cancer Epidemiol Biomarkers Prev.* **11,** 630–639.

Duffield-Lillico, A. J., Slate, E. H., Reid, M. E., Turnbull, B. W., Wilkins, P. A., Combs, G. F. J., Park, H. K., Gross, E. G., Graham, G. F., Stratton, M. S., Marshall, J. R., and Clark, L. C. (2003). Selenium supplementation and secondary prevention of nonmelanoma skin cancer in a randomized trial. *J Natl Cancer Inst* **95,** 1477–1481.

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* **68,** 589–599.

Elashoff, M. and Ryan, L. (2004). An em algorithm for estimating equations. *Journal of Computational and Graphical Statistics* **13,** 46–65.

Fan, J. and Li, R. (2002). Variable selection for cox's proportional hazards model and frailty model. *The Annals of Statistics* **30,** 74–99.

Field, C. A. and Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society, Series B* **69,** 369–390.

Fuchs, H. J., Borowitz, D. S., Christiansen, D. H., Morris, E. M., Nash, M. L., Ramsey, B. W., Rosenstein, B. J., Smith, A. L., and Wohl, M. E. (1994). Effect of aerosolized recombinant human dnase on exacerbations of respiratory symptoms

and on pulmonary function in patients with cystic fibrosis. the pulmozyme study group. *New England Journal of Medicine* **331,** 637–642.

Godambe, V. P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika* **63,** 277–284.

Godambe, V. P. (1991). *Estimating Functions*. Oxford University Press.

Goldstein, H. (1991). Nonlinear multilevel models, with an application to discrete response data. *Biometrika,* **78,** 45–51.

Goldstein, H. (1999). *Multilevel Statistical Models*.

Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* **159,** 505–513.

Ha, I. D. and Lee, Y. (2005). Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models. *Biometrika* **92,** 717–723.

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

Hall, D. B. (2001). On the application of extended quasi-likelihood to the clustered data case. *The Canadian Journal of Statistics* **29,**.

Huang, C.-Y. and Wang, M.-C. (2004). Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* **99,** 1153–1165.

Huang, Y. (2002). Censored regression with the multistate accelerated sojourn times model. *Journal of the Royal Statistical Society, Series B* **64,** 17–29.

Huang, Y. and Wang, M.-C. (2003). Frequency of recurrent events at failure time: Modeling and inference. *Journal of the American Statistical Association* **98,** 663–670.

Hunter, D. R. and Lange, K. (2003). A tutorial on mm algorithms.

Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *Journal of the American Statistical Association* **93,** 720–729.

Jorgensen, B. and Knudsen, S. J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics* **31,** 93–114.

Kaas, R., Dannenburg, D., and Goovaerts, M. (1997). Exact credibility for weighted observations. *ASTIN Bulletin* **27,** 287–295.

Kortram, R. A., van Rooij, A. C. M., Lenstra, A. J., and Ridder, G. (1995). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica* **49,** 269–281.

Kosorok, M. R., Lee, B. L., and Fine, J. P. (2004). Robust inference for univariate proportional hazards frailty regression models. *The Annals of Statistics* **32,** 1448–1491.

Kuk, A. Y. (1995). Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society. Series B (Methodological)* **57,** 395–407.

Lange, K. (1995). A gradient algorithm locally equivalent to the em algorithm. *Journal of the Royal Statistical Society, Series B* **57,** 435–437.

Lawless, J. F. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *The Canadian Journal of Statistics* **26,** 549–565.

Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly* **26,** 403–413.

Li, Y., Ryan, L., Bellamy, S., and Satten, G. (2003). Inference on clustered survival data using imputed frailties. *Journal of Computational and Graphical Statistics* **12,** 640–662.

Lin, D. Y., Wei, L. J., and Ying, Z. (2001). Semiparametric transformation models for point processes. *Journal of the American Statistical Association* **96,** 620–628.

Lin, X. and Breslow, N. E. (1996). Bias correction in generalised linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association* **91,** 1007–1016.

Ma, R. (1999). *An Orthodox BLUP approach to Generalized Linear Models*. PhD thesis, University of British Columbia.

Ma, R. and Jorgensen, B. (2001). Generalized linear mixed models: An orthodox blup approach. Technical report, University of New Brunswick.

Ma, R. and Jørgensen, B. (2007). Nested generalized linear mixed models: an orthodox best linear unbiased predictor approach. *Journal of the Royal Statistical Society, Series B* **69,** 625–641.

Ma, R., Krewski, D., and Burnett, R. T. (2003). Random effects cox models: A poisson modelling approach. *Biometrika* **90,** 157–169.

Ma, R., Willms, J. D., and Burnett, R. T. (2001). A nested frailty survival model for recurrent events. In Chaubey, Y. P., editor, *Recent Advances in Statistical Methods. Proceedings of Statistics 2001 Canada: The 4th Conference in Applied Statistics*, pages 186–197.

Murphy, S. A. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics* **23,** 182–198.

Ng, E. T. M. and Cook, R. J. (1997). Modeling two-state disease processes with random effects. *Lifetime Data Analysis* **3,** 315–335.

Ng, E. T. M. and Cook, R. J. (1999). Robust inference for bivariate point processes. *Canad. J. Statist.* **27,** 509–524.

Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorenson, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **12,** 25–43.

Nielsen, J. P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. *Scandinavian Journal of Statistics* **28,** 675–698.

Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer, New York.

Oakes, D. and Cui, L. (1994). On semiparametric inference for modulated renewal processes. *Biometrika* **81,** 83–90.

Ohlsson, E. (2005). Simplified estimation of structure parameters in hierarchical credibility. Research Report 5, Stockholm University Mathematical Statistics.

Ohlsson, E. and Johansson, B. (2003). Credibility theory and glm revisited. Research Report 15, Stockholm University Mathematical Statistics.

Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics* **26,** 183–214.

Peña, E. A., Strawderman, R. L., and Hollander, M. (2001). Nonparametric estimation with recurrent event data. *Journal of the American Statistical Association* **96,** 1299–1315.

Pierce, D. A., Stewart, W. H., and Kopecky, K. J. (1979). Distribution-free regression analysis of grouped survival data. *Biometrics* **35,** 785–793.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics* **22,** 2031–2050.

Prentice, R. L. and Gloeckle, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34,** 57–67.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68,** 373–379.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* **47,** 825–839.

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

Rodriguez, G. and Goldman, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* **158,** 73–89.

Schaubel, D. E. and Cai, J. (2005). Analysis of clustered recurrent event data with application to hospitalization rates among renal failure patients. *Biostatistics* **6,** 404–419.

Song, P. X.-K. and Jiang, W. (1998). Tweedie class log-linear models with random effects and internal correlation.

Strawderman, R. L. (2005). The accelerated gap times model. *Biometrika* **92,** 647–666.

Strawderman, R. L. (2006). A regression model for dependent gap times. *The International Journal of Biostatistics* **2,** Issue 1, Article 1. Available at: http://www.bepress.com/ijb/vol2/iss1/1.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer.

Therneau, T. M. and Hamilton, S. A. (1997). rhdnase as an example of recurrent event analysis. *Statistics in medicine* **16,** 2029–2047.

Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Statistics in Medicine* **16,** 385–395.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

Whitehead, J. (1980). Fitting cox's regression model to survival data using glim. *Applied Statistics* **29,** 268–275.

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

Xue, X. (1998). Multivariate survival data under bivariate frailty: An estimating equation approach. *Biometrics* **54,** 1631–1637.

Xue, X. and Brookmeyer, R. (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis* **2,** 277–289.

Yan, J. and Fine, J. P. (2005). Analysis of episodic data with application to recurrent pulmonary exacerbations in cystic fibrosis patients.

Yashin, A. I., Iachine, I. A., Begun, A. Z., and Vaupel, J. W. Hidden frailty: Myths and reality.

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

Ye, Y., Kalbfleisch, J. D., and Schaubel, D. E. (2007). Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* **63,** 78–87.