Lifetime Data Anal (2011) 17:3–28
DOI 10.1007/s10985-010-9162-0

Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches

Timothy E. Hanson · Adam J. Branscum · Wesley O. Johnson

Received: 1 September 2009 / Accepted: 15 March 2010 / Published online: 6 April 2010

Abstract The joint modeling of longitudinal and survival data has received extraordinary attention in the statistics literature recently, with models and methods becoming increasingly more complex. Most of these approaches pair a proportional hazards survival with longitudinal trajectory modeling through parametric or nonparametric specifications. In this paper we closely examine one data set previously analyzed using a two parameter parametric model for Mediterranean fruit fly (medfly) egg-laying trajectories paired with accelerated failure time and proportional hazards survival models. We consider parametric and nonparametric versions of these two models, as well as a proportional odds rate model paired with a wide variety of longitudinal trajectory assumptions reflecting the types of analyses seen in the literature. In addition to developing novel nonparametric Bayesian methods for joint models, we emphasize the importance of model selection from among joint *and* non joint models. The default in the literature is to omit at the outset non joint models from consideration. For the medfly data, a predictive diagnostic criterion suggests that both the choice of survival model and longitudinal assumptions can grossly affect model adequacy and prediction. Specifically for these data, the simple joint model used in by Tseng et al.

Electronic supplementary material The online version of this article (doi:10.1007/s10985-010-9162-0) contains supplementary material, which is available to authorized users.

T. E. Hanson (✉)
Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA
e-mail: hans0058@umn.edu

A. J. Branscum
Departments of Biostatistics, Statistics, and Epidemiology, University of Kentucky, Lexington, KY 40536, USA

W. O. Johnson
Department of Statistics, University of California, Irvine, Irvine, CA 92697, USA

(Biometrika 92:587–603, 2005) and models with much more flexibility in their longitudinal components are predictively outperformed by simpler analyses. This case study underscores the need for data analysts to compare on the basis of predictive performance different joint models and to include non joint models in the pool of candidates under consideration.

Keywords Mixture of Polya trees · Model selection · Predictive inference · Survival analysis · Time dependent covariates

1 Introduction

Studies designed to relate event times to covariates often involve the incorporation of measurements on one or more longitudinal processes (e.g. repeated measurements taken on biomarkers or exposure histories) that are thought to be associated with the time to event occurrence. Statistical models for such data have been termed "joint longitudinal/survival" models and in this paper they are referred to as joint models. In general, the primary inferential objectives of joint modeling are to characterize trends in the time course of relevant longitudinal processes, to determine which covariate processes or variables are predictive of time to event occurrence, and to assess the magnitude of such associations, while adjusting for confounder processes and/or covariables. We subsequently use the terminology longitudinal process, covariate process, and time dependent covariate (TDC), interchangeably.

Traditionally, these objectives were achieved by disjoint analysis of the longitudinal data (e.g. using mixed models or other standard analytic methods for temporal data to obtain point estimates of longitudinal trajectories) and the event time data (e.g. the Cox model with known TDC's). However, conducting a separate longitudinal data analysis may provide biased inferences in the Cox model when a TDC is associated with event occurrence, resulting in longitudinal data that are subject to informative missingness. Moreover, a separate or raw survival analysis that simply conditions on a TDC $y(t)$ can also produce biased parameter estimates when the "true" process, say $x(t)$, is measured with error or is subject to random biological variability (Prentice 1982; Bycott and Taylor 1998). In the absence of measurement error the process $y(t) = x(t)$ is observed. Sometimes the longitudinal process can be accurately characterized by a piecewise constant function with changes at discrete time points; Hanson et al. (2009) considered several flexible Bayesian semiparametric models in this context. However, when event times are modeled as dependent on a continually varying, but finitely sampled $x(t)$ only through the last observed value, this approach is akin to last observation carried forward (LOCF) imputation, which can also be prone to biased parameter estimates (Prentice 1982).

Alternatively, two-stage procedures attempt to improve upon the simplicity of LOCF methods by imputing unobserved values of $x(t)$ by first modeling the longitudinal process and then treating the imputed trend as a known TDC in a survival model. This approach has the potential for biased estimates when $x(t)$ is informatively censored at the event time. Although the introduction of estimation bias in Cox

model regression coefficients has been well-documented in the literature, it is unclear whether naive approaches such as LOCF reduce the *predictive* ability of a model.

Drawbacks of separate analyses and two-stage procedures motivated the recent flourish of research on joint models for longitudinal and survival data (see Tsiatis and Davidian 2004 for a review). Joint modeling directly accounts for the relationship between event time and the implicit censoring of the corresponding repeated measures sequence. Moreover, it is expected that, with joint modeling, there is opportunity for increased statistical efficiency due to using all of the data simultaneously in a single model, as well as appropriate assessment of estimator variability obtained by correctly treating the longitudinal process as random rather than incorrectly assuming there is no uncertainty attached to imputed values from a separate longitudinal model.

Joint modeling links the longitudinal and survival data by factoring the joint likelihood into a conditional survival component in which event times are modeled to be dependent on a latent process $x(t)$, which is itself modeled appropriately. There are now several published Bayesian approaches to joint analysis, including methods developed by Faucett and Thomas (1996), Wang and Taylor (2001), and Brown and Ibrahim (2003). Frequentist approaches were developed by Wulfsohn and Tsiatis (1997), Song et al. (2002), and Law et al. (2002), among others. Ibrahim et al. (2001, Chap. 7) and Tsiatis and Davidian (2004) provide overviews of research on statistical methodology for joint longitudinal/survival models. The initial development of joint models was largely motivated by data from AIDS clinical trials that were designed to evaluate the therapeutic effects of treatments on the development of AIDS or death, where CD4 count and viral loads were used as markers for disease escalation (see, for example, Wang and Taylor 2001 and references therein).

In this paper we develop several Bayesian semiparametric joint models that achieve flexibility through arbitrary baseline survival functions, which are modeled with mixtures of finite Polya trees (MFPT) priors (Lavine 1992; Hanson and Johnson 2002; Hanson 2006). This approach generalizes standard parametric survival regression analysis, including log-normal, Weibull, and log-logistic regression. The survival models we consider include the Cox (1972) model, a generalization of the proportional odds (PO) model (Sundaram 2006), and a generalization of the accelerated failure time (AFT) model (Cox and Oakes 1984), which we refer to as the CO model. Details about these models are provided in Sect. 2.

A prominent feature that sets the current study apart from other published joint modeling applications is our detailed treatment of model comparison according to prediction accuracy. We compare models (Cox, PO and CO) using the predictive approach of Geisser and Eddy (1979), which employs conditional predictive ordinates in its calculation. They propose what they call the pseudo marginal likelihood, which is an approximation to the usual marginal likelihood (sometimes called the prior predictive) that is used in calculating Bayes factors. The ratio of pseudo marginal likelihoods corresponding to two distinct models, for instance the PO and CO, is called a pseudo Bayes factor for comparing these models. We are not aware of any other papers that have made such comparisons across semiparametric families for joint modeling. Instead, generally, a single joint model is developed without regard to model selection or fit. We also use pseudo marginal likelihoods to compare raw versus two-stage versus

joint modeling to decide if, for a given data set, joint or two stage methods predictively outperform a simpler raw analysis. We also compare semiparametric models to their parametric counterparts.

Model development for the three semiparametric approaches is presented in the next section. In Sect. 3 we detail Bayesian semiparametric methods for joint modeling, discuss approaches to Markov chain Monte Carlo (MCMC) simulation to enable posterior and predictive inference, and develop predictive model selection in this context. Novel aspects of our work include extending the CO model to a Bayesian joint specification, extending the PO model to a joint specification, and comparison of competing joint models from a predictive standpoint. Of particular interest to us is the recent development of frequentist joint modeling methodology for the CO semiparametric survival model in conjunction with longitudinal data by Tseng et al. (2005). Their methodology was applied to an analysis of female Mediterranean fruit fly (called "medfly" for short) lifetimes modeled as a function of reproductive fertility, which was measured daily over the lifespan by the number of eggs laid per day. They implemented a Monte Carlo EM algorithm to obtain a point estimate of the regression coefficient in conjunction with a bootstrap standard error. We follow up with an analysis of these data in Sect. 4, only with a direct comparison of the Cox, CO and PO models. Our methods allow for broader inferences since, due to MCMC methodology, it is simple to make inferences about any quantity of interest, for example, a hazard function, or a ratio of hazard functions, if desired. With regard to our model selection criterion, we establish the futility of joint or two-stage modeling for these data. Concluding remarks are provided in Sect. 5.

2 Model development

Consider the following situation. One or more longitudinal processes are observed on individual i until either the individual experiences an event of interest, leaves the study, or the study is terminated. The latter two scenarios correspond to censoring. Censoring mechanisms are assumed independent of event times. For simplicity of exposition, we only consider a single covariate process since there is no conceptual or practical difficulty in our approach associated with having additional ones. Thus, consider a single longitudinal process for the ith experimental unit, $x_i(t)$, that is measured with non-negligible error so that the observed process is $y_i(t) = x_i(t) + \epsilon_i(t), t > 0$. The full longitudinal model is developed later in this section. Note that time dependent covariates that are not observed with error are also readily handled within the model (Hanson et al. 2009).

Next, let T_i denote either the event time or the censoring time, whichever is observed first, and let δ_i be an indicator of event occurrence. We subsequently model the conditional for $T_i \,|\, x_i(\cdot)$ according to Cox, CO and PO models. Since the process $y_i(\cdot)$ is only observed at a fixed vector of times, say $\mathbf{t}_i = (t_{i1}, \ldots, t_{im_i})$, we define the observed value to be the vector $\mathbf{y}_i = (y_i(t_{i1}), \ldots, y_i(t_{im_i}))$. The time vectors will generally vary from individual to individual. In our application, the trajectory $x_i(t)$ will be described fully by a finite-dimensional vector $\mathbf{b}_i$ in a random coefficient regression model with precision τ. The model for the data can be represented generically as a joint probability density

$$f(\mathbf{y}_i, \mathbf{b}_i, T_i|\tau, \boldsymbol{\pi}) = f(\mathbf{y}_i|\mathbf{b}_i, \tau) f(\mathbf{b}_i|\boldsymbol{\pi}_l) f(T_i|\mathbf{b}_i, \boldsymbol{\pi}_s).$$

Here, $\boldsymbol{\pi}_l$ is the parameter vector corresponding to the longitudinal component of the model, and $\boldsymbol{\pi}_s$ is comprised of all parameters associated with the semiparametric survival component. The likelihood contribution for a non-censored individual is given by the integral of the above joint density with respect to $\mathbf{b}_i$. For the models we consider it is not necessary to do the integration explicitly.

To summarize, the data for n independently sampled subjects are denoted by $\{(T_i, \delta_i, \mathbf{y}_i, \mathbf{z}_i, \mathbf{t}_i)\}_{i=1}^n$. The vector $\mathbf{y}_i$ contains observed sequential values of the process $y_i(\cdot)$, which is subject to inherent biological variability and/or measured with error at times $\mathbf{t}_i$. The vector $\mathbf{z}_i$ denotes the collection of baseline covariates, thus their values are known and fixed through time. We assume that events and censoring are independent of the process measurement schedule. Assessment and quantification of the association between the potentially latent biological process $x_i(t)$ and event time T_i is the primary inferential objective.

2.1 Survival component

Several regression models associating survival time with a covariate process have been proposed in the literature, including models due to Cox (1972), Prentice and Kalbfleisch (1979), Aalen (1980), Cox and Oakes (1984), and Sundaram (2006). We consider Bayesian joint analysis with three of these models. Define $X_t = \{x(s) : s \leq t\}$ to be the history of the process $x(\cdot)$ up to time t, and let $h_0(t)$ and $S_0(t) = \exp\{-\int_0^t h_0(s)ds\}$ denote the baseline hazard and survival functions, respectively. Finally, let the parameter β represent the regression effect of the true process on event time. The Cox (1972) model as defined through the hazard function is

$$h(t|X_t) = e^{x(t)\beta} h_0(t). \tag{1}$$

The generalization of the accelerated failure time model due to Cox and Oakes (1984) is defined by

$$S(t|X_t) = S_0\left(\int_0^t e^{x(s)\beta} ds\right), \tag{2}$$

and the generalization of the proportional odds model due to Sundaram (2006) is given by

$$\frac{d}{dt}\left[\frac{1 - S(t|X_t)}{S(t|X_t)}\right] = e^{x(t)\beta} \frac{d}{dt}\left[\frac{1 - S_0(t)}{S_0(t)}\right]. \tag{3}$$

There is no published work that we are aware of involving joint model specifications for (3) or Bayesian joint models for (2). The Bayesian joint models we develop here extend the work of Hanson et al. (2009), who presented Bayesian semiparametric methods with fixed TDC's for the Cox and CO models, and for another generalization of the AFT model due to Prentice and Kalbfleisch (1979).

The vast majority of Bayesian and frequentist approaches to joint modeling use the Cox model defined in (1) for the survival component. This model, however, incorporates a potentially unrealistic assumption for time-varying predictors, namely that the risk of event at time t depends only on the current value of $x(t)$ and not on its history. Although this assumption might be valid in some cases, a cumulative effect of exposure or treatment history will be biologically appropriate in other cases. To address this limitation, Cox and Oakes (1984) generalized the AFT model by relating the hazard of failure at time t to a measure of the "average" time-dependent covariate effect up to time t, defined by $\bar{c}(t) = t^{-1} \int_0^t e^{x(s)\beta} \mathrm{d}s$. The following representation of the CO model indicates that an individual with TDC $x(\cdot)$ exhausts their lifetime at a rate of $e^{x(t)\beta}$ relative to their (counterfactual) baseline rate:

$$T_0 = \int_0^T e^{x(s)\beta} \mathrm{d}s, \quad T_0 \sim S_0.$$

The corresponding hazard function is $h(t|X_t) = e^{x(t)\beta} h_0(\bar{c}(t)t)$, with survival function defined in (2). The hazard function reflects both immediate ($e^{x(t)\beta}$) and cumulative ($\bar{c}(t)$) effects of the covariate process $x(\cdot)$ on risk. Note that the CO hazard reduces to $h(t|x) = e^{x\beta} h_0(e^{x\beta}t)$, the hazard for a standard AFT model, when the covariate does not vary over time ($x(t) \equiv x$).

Our Bayesian analysis requires a prior probability model for $S_0(\cdot)$ and β. We discuss a nonparametric prior for $S_0(\cdot)$ in Sect. 2.3. Our model will assume independence of these parts, and we use an improper uniform prior for β. However, because of our prior on the baseline survivor function, it is straightforward to incorporate the informative priors for β that are discussed in Bedrick et al. (2000) for fixed (or averaged) covariates.

2.2 Longitudinal component

Laird and Ware (1982) popularized the linear mixed model, a tool especially useful for modeling longitudinal data. A generalization of their model for the process associated with subject i is

$$y_i(t) = x_i(t) + \epsilon_i(t); \quad x_i(t) = f(t) + g_i(t) + \mathbf{z}_i'\boldsymbol{\alpha} + \mathbf{v}_i'\mathbf{c}_i, \tag{4}$$

where $\epsilon_i(t) \overset{iid}{\sim} N(0, 1/\tau)$ are random and independent normal errors, $f(t)$ is a fixed function of time and $g_i(t)$ represents random deviation from f for individual i. The $\mathbf{c}_i$'s are random effects, $\mathbf{z}_i$ and $\mathbf{v}_i$ are design/baseline covariate vectors, and $\boldsymbol{\alpha}$ is a vector of regression coefficients. The functions $f(\cdot)$ and/or $g_i(\cdot)$ have been modeled in joint analysis applications with polynomials (Wang and Taylor 2001; Brown and Ibrahim 2003), Gaussian processes (Wang and Taylor 2001), and B-splines (Brown et al. 2005).

It is common in longitudinal data analysis to represent $f(\cdot)$ and $g_i(\cdot)$ in terms of finite basis function expansions (e.g. polynomials, fractional polynomials, splines,

wavelets, Fourier expansions, et cetera) and to thus write each as a linear combination, e.g. $f(t) = \sum_{k=1}^{K} \mu_k \psi_k(t)$ and $g_i(t) = \sum_{k=1}^{K} b_{ik} \psi_k(t)$ with coefficients $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)'$ for f and random effects $\mathbf{b}_i = (b_{i1}, \ldots, b_{iK})'$ for g_i. In these instances, model (4) can be represented as a linear mixed model. Standard distributional assumptions posit the vectors $\mathbf{b}_i$ as iid multivariate normal with mean $\mathbf{0}$ and unknown covariance matrix $\boldsymbol{\Sigma}$. In the case of a penalized spline, there is an additional smoothing parameter that can be viewed as a variance component (Li et al. 2009). The number and placement of knots can also be regarded as smoothing parameters. Random effects, $\mathbf{c}_i$, are generally modeled with multivariate normal distributions with mean zero and an unknown covariance.

The mixed model can approximate a Gaussian process by taking $\psi_1(\cdot), \ldots, \psi_K(\cdot)$ to be kernel functions over a set of equispaced knots, much like B-splines. For example, assuming $\mathbf{b}_i \sim N_K(\mathbf{0}, \sigma^2 \mathbf{I})$, the process

$$x(t) = \sum_{k=1}^{K} b_{ik} \phi\{(t - l_k)/\kappa\}/\kappa$$

where $\phi(x) = \exp(-0.5x^2)$ and $\mathbf{l} = (l_1, \ldots, l_K)$ are knots, approximates a Gaussian process with a Gaussian covariance function (Higdon 2001). The covariance depends on (σ, κ); the full conditional for κ is typically sampled using a Metropolis-Hastings step, or the prior for κ can be discrete on a grid of reasonable values relative to $\Delta = l_{k+1} - l_k$. Unlike partially sampling a Gaussian process over a mesh, the convolution approach obviates taking the inverse of a large dimensional matrix at each MCMC iteration. Thus, this approach is able to "... write the stochastic process ... as a 'random effects model' and thereby develop new joint model implementations ..." as suggested by Tsiatis and Davidian (2004) as an avenue of unexplored research. See also Kneib (2006) and Banerjee et al. (2008) for related ideas. We consider this approach in Sect. 4.

Our focus in the current study is use of joint modeling for making appropriate inferences about the effect of a longitudinal process on survival. While we do allow for considerable generality in modeling the longitudinal part along with associated detail, it is beyond the scope of this article to attempt to cover the myriad of possible implementations beyond what we consider here. There are many possible choices and going into these issues would distract from our main purpose. Our illustrations give details including prior specification for the specific longitudinal models employed herein.

2.3 Prior for baseline survival

All three semiparametric models that were introduced above in (1), (2), and (3) involve a baseline survivor function $S_0(\cdot)$. Our approach is to place competing survival models on "equal ground" in terms of prior specification, partly for parsimony and partly since we plan to compare these models. We thus propose to place a single nonparametric prior on S_0 for all three models.

We have chosen to model $S_0(\cdot)$ with a (finite) mixture of Polya trees prior in each instance because, in addition to other attractive properties (Hanson 2006), with this prior it is possible to embed a parametric family of survivor functions within a broader class of survival models defined by it. This is termed centering on the parametric family since the prior expectation of $S_0(\cdot)$ includes all survival functions from the specified family. In our illustrations, all three baseline survivor functions are modeled with a single MFPT prior that has a log-logistic centering family. In addition to representing a nonparametric generalization of certain standard parametric survival models, a further benefit is the relative ease of implementing MCMC computing algorithms for model fitting (e.g. Hanson and Yang 2007; Branscum and Hanson 2008) that comes from directly making use of output from parametric analyses in constructing efficient proposal distributions for Metropolis steps within Gibbs samplers. Inferences for survival and hazard functions based on an MFPT analysis are directly available because S_0 is not marginalized; it is sampled at each MCMC iteration.

Foundational work on Polya trees was presented by Ferguson (1974), Lavine (1992), Lavine (1994), and Mauldin et al. (1992). Mixtures of Polya trees were developed by Berger and Guglielmi (2001), Hanson and Johnson (2002), and Hanson (2006). See also Paddock et al. (2003) for related work. Briefly, a realization of a simple Polya tree is determined by (i) a tree that has been created through the successive binary partitioning of sets in the sample space, and (ii) a corresponding tree of conditional probabilities. The sample space for survival analysis corresponds to $[0, \infty)$; the first split on the tree corresponds to a number in this set. The first "branch" probabilities are $S_0(B_0)$ and $S_0(B_1)$, where $[0, \infty) = B_0 \cup B_1$; these probabilities must add to one. The second level of the tree involves splitting B_0 and B_1, each into two additional mutually exclusive and exhaustive sets. We similarly define the next set of branch probabilities as $S_0(B_{00}|B_0)$, $S_0(B_{01}|B_0)$, and $S_0(B_{10}|B_1)$, $S_0(B_{11}|B_1)$, where each pair also add to one. The next level entails splitting each of these sets, and so on ad infinitum. If a member of each pair of conditional branch probabilities is modeled independently with its own beta distribution, we have constructed a (random) Polya tree that places prior probability on a broad class of probability distributions.

Suppose the centering family is $\{G_{\boldsymbol{\theta}}(\cdot) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$. Then, if the beta priors on the branch probabilities have mean 0.5, and if, for fixed $\boldsymbol{\theta}$, the mth level partition of the tree is determined by the $1/2^m$ quantiles of $G_{\boldsymbol{\theta}}(\cdot)$, it follows that $E(S_0(\cdot)) = G_{\boldsymbol{\theta}}(\cdot)$. Finally, a common choice for the beta distribution at level m of the tree is beta(cm^2, cm^2), where c is a positive constant. With this choice, we say that $S_0(\cdot) \sim PT(c, G_{\boldsymbol{\theta}}(\cdot))$. The resulting Polya tree will be absolutely continuous with probability one. In addition, if c is large, then the random $S_0(\cdot)$ is approximately equal to the fixed $G_{\boldsymbol{\theta}}(\cdot)$; a small value of c provides for greater flexibility.

Hanson (2006) argues that a good approximation to a full mixture of Polya trees involves truncation at level M, where M is selected so that $2^M \approx n/N$ and N is a typical number falling into each set at level M. Selecting M in this manner is based on the premise that going beyond this level would yield very weakly identified conditional probabilities on the highest levels. For fixed $\boldsymbol{\theta}$, the corresponding prior is denoted as a finite Polya tree. Define the collection of pairs of branch probabilities up to level

M as $\mathcal{X}_M = \left\{(X_{\mathbf{e}0}, X_{\mathbf{e}1}) : \mathbf{e} \in \{0,1\}^{M-1}\right\}$. For example, $X_{101} = S_0(B_{101}|B_{10})$. A mixture of finite Polya trees prior for S_0 obtains by placing a prior on $\boldsymbol{\theta}$ and/or c. This is hierarchically written $S_0|c, \boldsymbol{\theta} \sim FPT(c, G_{\boldsymbol{\theta}})$, $(c, \boldsymbol{\theta}) \sim p(\boldsymbol{\theta})p(c)$, which implies for any binary number $\mathbf{e} \in \{0,1\}^{M-1}$ that at level $m < M$, $X_{\mathbf{e}0} \sim \text{beta}(cm^2, cm^2)$, i.e. $(X_{\mathbf{e}0}, X_{\mathbf{e}1}) \sim \text{Dirichlet}(cm^2, cm^2)$. The prior is constructed so that sets beyond level M have S_0-probability coinciding with $G_{\boldsymbol{\theta}}$. With this construction, the centering family can be chosen from among commonly used parametric models such as lognormal, log-logistic, or Weibull. The log-logistic, which we subsequently use in Sect. 4, is indexed by $\boldsymbol{\theta} = (\theta_1, \theta_2)$, with

$$G_{\boldsymbol{\theta}}(t) = \frac{1}{1 + t^{1/\theta_2} e^{-\theta_1/\theta_2}}.$$

The MFPT model is ultimately not nonparametric due to the truncation, but rather is a richly parametric model because of having a total of $2^M - 1$ branch probabilities. The nature of a nonparametric MFPT analysis can be visualized as an underlying parametric family $G_{\boldsymbol{\theta}}$ capturing overall data shape and trend, refined by conditional branch probabilities that provide latitude for local, data-driven fluctuations about the parametric family.

3 Inference, implementation, and model choice

In this section we discuss some of the types of inferences imagined for this setting, some details about MCMC sampling from the joint posterior, which includes sampling the latent processes $x_i(t)$, under the joint model, and methods for predictive comparison of models (1–3) for survival. We also ultimately compare models based on traditional TDC's versus imputation of the longitudinal process versus joint analysis.

3.1 Inference

Standard inferences will involve the usual point and interval estimates for the regression parameter β, although β is only interpretable relative to baseline for the CO model. We are also interested in estimating the predictive lifetime density for individuals with given longitudinal trajectories. For raw (LOCF) data we assume $x_i(t) = 0$ for $t > T_i$; alternatives would be to set $x_i(t)$ at the last value recorded or impute values using time-series methods. For this reason, the density estimates from the raw analysis are only known up to the last observation time before death. Beyond this value, there is a fixed amount of probability mass left, but it is unknown how that mass will be spread out because the (internal) trajectory vanishes.

For two-stage analysis based on a longitudinal model of the form $x_i(t) = \sum_{k=1}^{K} b_{ik}\psi_k(t)$, we use the posterior mean $\hat{x}_i(t) = E\{x_i(t|\mathbf{b}_i)|\mathbf{y}_1, \ldots, \mathbf{y}_n\} = \sum_{k=1}^{K} \hat{b}_{ik}\psi_k(t)$ based on the longitudinal model only for $\mathbf{b}_i$; here $\widehat{\mathbf{b}}_i = E\{\mathbf{b}_i|\mathbf{y}_1, \ldots, \mathbf{y}_n\}$. For the medfly data, these estimates are remarkably close to the posterior Bayes estimates from the full joint models (Figs. 1, 2, 3, 4, 5, 6), and we see no loss of

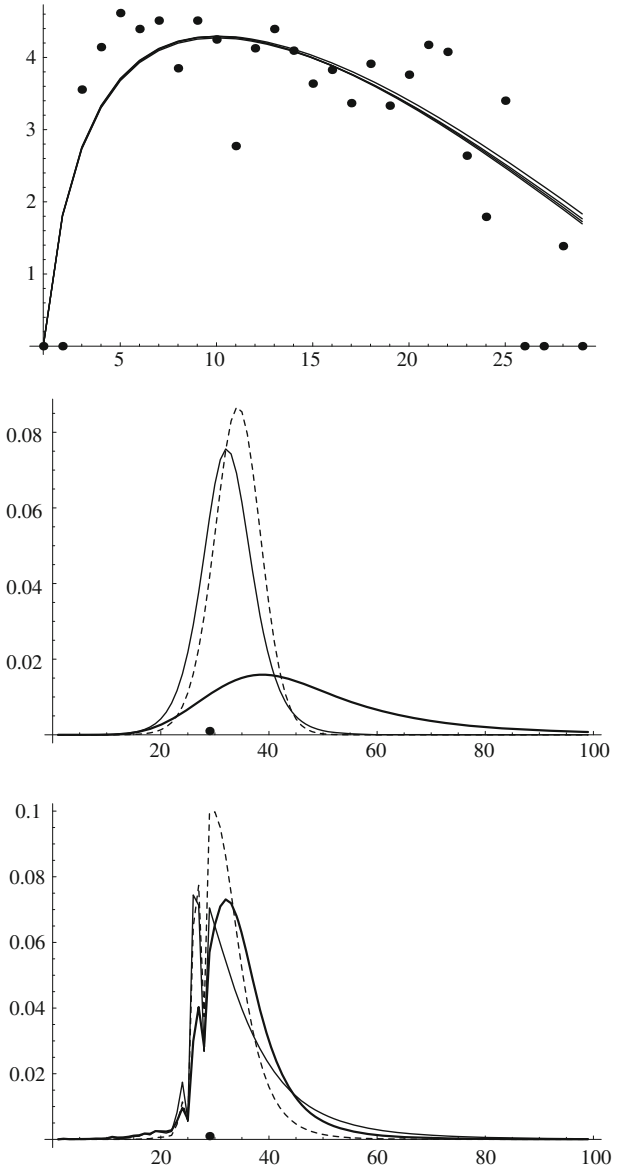

Fig. 1 Fly 247 (event time 29 days). *Top panel* is raw and estimated trajectories, *second panel* is predictive densities from parametric imputed analyses, and *bottom panel* is predictive densities from parametric LOCF analyses. PO, Cox, and CO (*Thin solid*, *dashed*, and *thick solid*, respectively)

information in this much more tractable estimation method. As $x(\cdot)$ is completely defined given $\mathbf{b}$ under the joint model, the predictive density is:

$$f(t|\mathbf{b}) = \int f(t|\mathbf{b}, \boldsymbol{\pi}_s)\, p(\boldsymbol{\pi}_s|\mathbf{y}_1, \ldots, \mathbf{y}_n, T_1, \ldots, T_n) d\boldsymbol{\pi}_s.$$

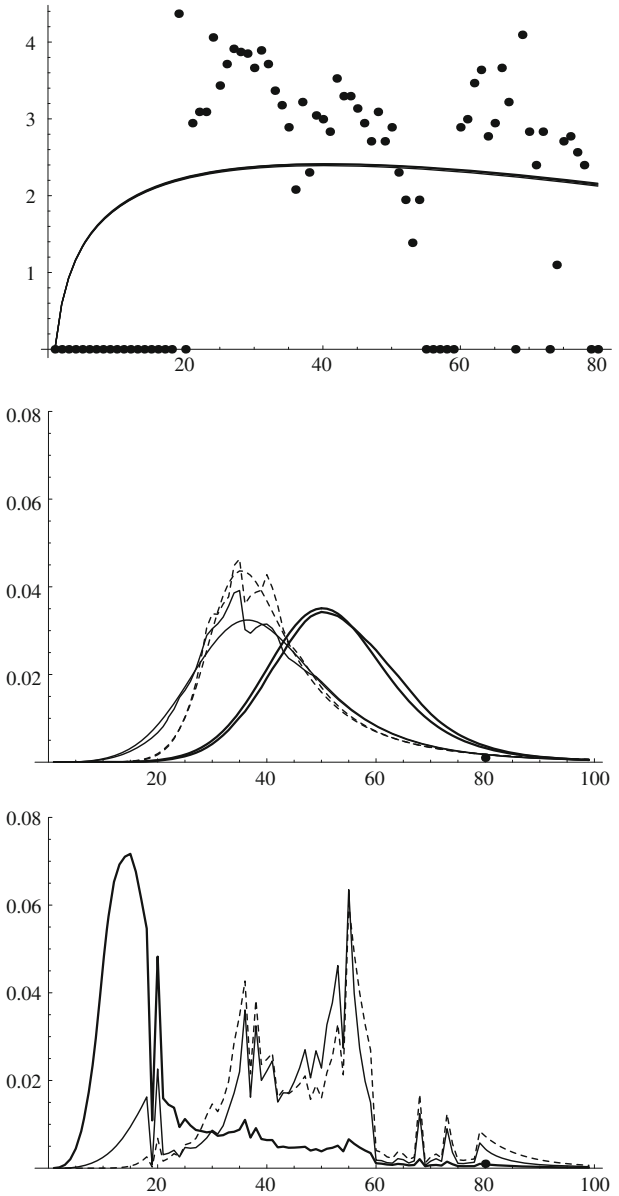

Fig. 2 Fly 22 (event time 80 days). *Top panel* are raw and estimated trajectories, *second panel* are predictive densities from parametric and MFPT imputed analyses, and *bottom panel* is predictive densities from MFPT LOCF analyses. PO, Cox, and CO (*Thin solid*, *dashed*, and *thick solid*, respectively)

Our final aim is to compare models using a predictive model selection criterion. The MCMC samples obtained from the full conditionals discussed in the next subsection make it possible to obtain numerical approximations for predictive inference. Details are in Sect. 3.3.

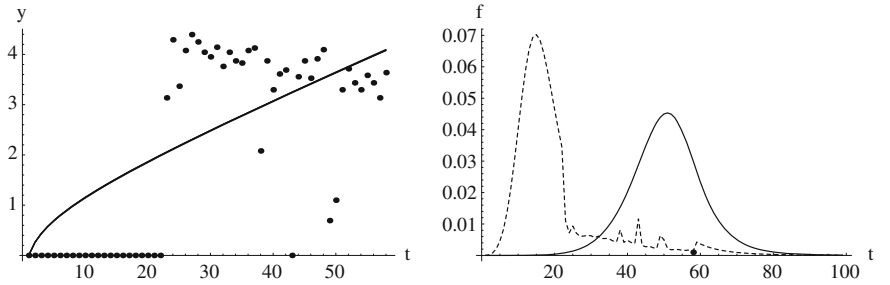

Fig. 3 Fly 38 (event time 58 days). *Left panel* is raw and estimated trajectories, *right panel* is predictive densities from parametric CO analyses. Imputed (*solid*) and raw (*dashed*)

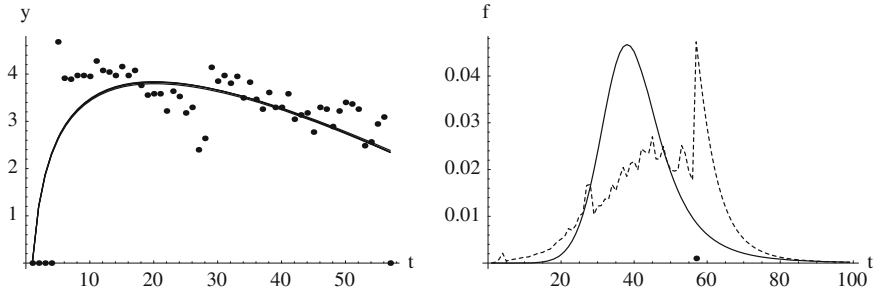

Fig. 4 Fly 45 (event time 57 days). *Left panel* is raw and estimated trajectories, *right panel* is predictive densities from parametric CO analyses. Imputed (*solid*) and raw (*dashed*)

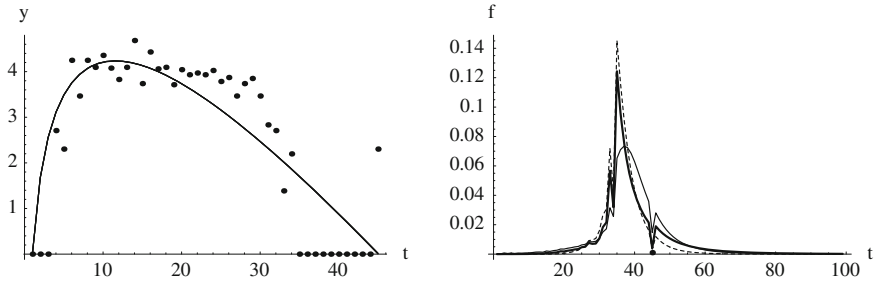

Fig. 5 Fly 6 (event time 45 days). *Left panel* is raw and estimated trajectories, *right panel* is predictive densities from parametric raw CO, Cox, and PO analyses (*thin solid*, *dashed*, and *thick solid*, respectively)

3.2 MCMC implementation

Under our model assumptions, including an MFPT prior for the baseline survivor distribution, the likelihood augmented with $\mathbf{b}_1, \ldots, \mathbf{b}_n$ is

$$\prod_{i=1}^{n} f(T_i|\mathbf{b}_i, \boldsymbol{\pi}_s)^{\delta_i} S(T_i|\mathbf{b}_i, \boldsymbol{\pi}_s)^{1-\delta_i} f(\mathbf{y}_i|\mathbf{b}_i, \tau) f(\mathbf{b}_i|\boldsymbol{\pi}_l).$$

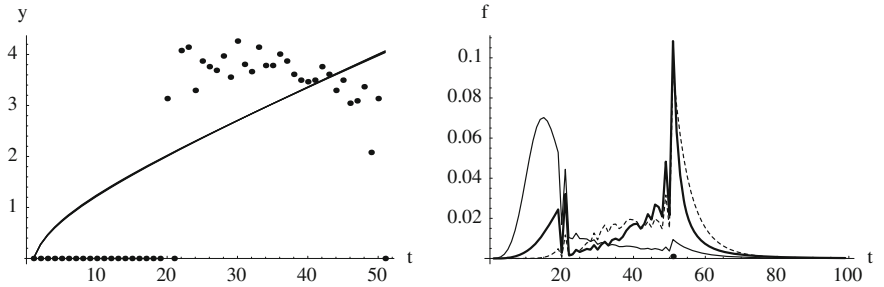

Fig. 6 Fly 111 (event time 51 days). *Left panel* is raw and estimated trajectories, *right panel* is predictive densities from parametric raw CO, Cox, and PO analyses (*thin solid*, *dashed*, and *thick solid*, respectively)

The survival parameters are $\boldsymbol{\pi}_s = (\beta, \boldsymbol{\theta}, \mathcal{X}_M)$, the longitudinal parameters are $\boldsymbol{\pi}_l = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and τ is the precision of $y_i(t_{ij})$. The pdf's, $f(T_i|\mathbf{b}_i, \boldsymbol{\pi}_s)$, and survivor functions, $S(T_i|\mathbf{b}_i, \boldsymbol{\pi}_s)$, were discussed by Hanson et al. (2009) for the Cox and CO models, and are developed in Appendix A for the PO model.

In our applications the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are updated from their full conditional normal and inverted Wishart distributions. The $\mathbf{b}_i$ are updated independently using either independence or random walk Metropolis-Hastings steps based on the full conditionals under the longitudinal model. Much more general longitudinal processes $x_i(t)$, such as Gaussian processes and B-splines, can also be accommodated. Details are given in Appendix B.

3.3 Model choice

Following theoretical considerations presented in Prentice (1982), Bycott and Taylor (1998) used simulations to show that LOCF and two-stage imputation by fitting a mixed model can lead to biased parameter estimates. However, when regression models are built as tools for characterizing future patient prognosis as they often are in joint analysis, selection criteria should target the identification of a predictively viable model. We consider model selection according to prediction accuracy by comparing parametric and semiparametric joint analysis to survival analysis with fixed TDC's and to two-stage models.

The cross-validated, pseudo marginal likelihood criterion (Geisser and Eddy 1979) is used here to quantify a model's ability to predict survival, $\prod_{i=1}^{n} p(T_i|\mathbf{T}_{-i}, \mathbf{y}_{1:n})$. On the log scale this becomes

$$\text{LPML} = \sum_{i=1}^{n} \log(\text{CPO}_i)$$

where $\text{CPO}_i = p(T_i|\mathbf{T}_{-i}, \mathbf{y}_{1:n})$ denotes the conditional predictive ordinate of event time T_i based on the remaining $\mathbf{T}_{-i} = \{T_j : j \neq i\}$ and the full longitudinal data $\mathbf{y}_{1:n} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$. The CPO_i statistic measures the amount of support the observed trajectory $\mathbf{y}_i$ gives the observed survival time $T_i = t_i$ through the model and the

remaining medflies. For predictive optimality the goal is to maximize LPML, and therefore, among the models under consideration, the one with the largest LPML is deemed superior. A corresponding pseudo Bayes factor comparing two models can be calculated as the ratio of their pseudo marginal likelihoods. Note that $p(T_i|\mathbf{T}_{-i}, \mathbf{y}_{1:n})$ is computed as either the pdf or the survival function depending on whether T_i is an event time or a censoring time.

The predictive LPML criterion has been recently promoted for use in Bayesian nonparametric model selection in part because it is straightforward to calculate CPO statistics from MCMC output. An identity tailored to the joint modeling setting is presented in Appendix C. Either within a given model or in comparing competing models, CPO statistics can highlight individuals with longitudinal profiles that are highly predictive of survival. We make extensive use of CPO statistics in contrasting the PO, Cox, and CO models in Sect. 4.

We stress that it would be inappropriate for us to use an LPML that was based on the joint distribution of longitudinal and survival time data since our inferential goal is completely focused on survival analysis. Joint modeling serves the purpose of handling time dependent processes appropriately. Moreover, we also compare models based on raw trajectories, making our choice of LPML all the more clear. An alternative method of (frequentist) prediction based model selection in survival analysis with TDCs was developed in Schoop et al. (2008).

4 Longevity of medflies

The data used for illustration came from a study reported in Carey et al. (1998) and further analyzed by Chiou et al. (2003) where the reproductive patterns of 1000 Mediterranean fruit flies were obtained by recording the number of eggs produced each day throughout their lifespan. The scientific goal was to examine the association between temporal trend in egg production and lifetime. A frequentist joint modeling approach based on the CO model (2) was used to analyze the most prolific egg-layers by Tseng et al. (2005), and like these authors we excluded from our analysis the flies whose lifetime production was less than 1145 eggs. We also removed the first two days from each trajectory, which have zero counts for all flies, to correspond to Fig. 2 in Tseng et al. (2005). This gave a sample size of 251 flies with lifetimes ranging from 22 to 99 days (24 to 101 days in the original data).

A survival analysis that treated egg production as a known TDC was implemented using a piecewise constant function with days representing changepoints. For comparison with the analysis by Tseng et al. (2005), we use the same longitudinal model they did (also used in Hsieh et al. 2006), and we also consider more flexible alternatives. They noted that scatterplots of temporal profiles of egg production suggested the nonlinear form that corresponds to a gamma kernel, namely $t^{b_1}e^{b_2(t-1)}$, with different b_1 and b_2 for each medfly. They transformed this nonlinear structure into a linear mixed model by considering $y_i(t) = \log\{N_i(t) + 1\}$, the natural log of one plus the number of eggs laid on day t. With this transformed longitudinal response, the trajectories are modeled as

$$y_i(t)|\mathbf{b}_i, \tau \sim N\left(b_{i1}\log t + b_{i2}(t-1), \tau^{-1}\right), \quad \mathbf{b}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma} \overset{iid}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{5}$$

Since there are no additional covariates for survival, a single regression coefficient β connects the survival model to the longitudinal process $x_i(t) = b_{i1}\log t + b_{i2}(t-1)$. The MFPT models used here set $M = 4$ and $c = 1$, with flat priors otherwise. About 16 observations fall into each of the 16 sets at level $M = 4$ if the log-logistic family is approximately correct. We also considered the prior $c \sim \Gamma(5, 1)$ for a subset of models, obtaining LPML values slightly smaller than with fixed $c = 1$.

Posterior propriety of the full models under a flat or Jeffreys' prior is not immediately obvious. However, Theorem 4.4 (p. 296) in Yang and Chen (1995), and results in Chen et al. (2002) indicate that posterior propriety for the longitudinal model under flat priors holds, at least for fixed τ. The longitudinal posterior is also proper under various reference priors. Fahrmeir and Kneib (2009) discuss posterior propriety for the Cox model with additive regression effects modeled via penalized B-splines; see also Chen et al. (2006). Establishing full posterior propriety under the additional flat prior $p(\beta) \propto 1$ could be done separately for each model, but we are unconcerned with this for our analyses since placing vague but proper priors produces essentially identical results for the medfly data. This is likely due to a relatively large amount of data going into each fly's trajectory, and that the survival times are uncensored.

4.1 Broad comparison across models

First, we fitted the models (1–3) with both the MFPT ($c = 1$) and parametric log-logistic model ($c \to \infty$). Each model was used in conjunction with: (i) raw trajectories only; (ii) the two-stage approach where longitudinal profiles are estimated from a mixed model, and then a survival analysis is performed treating the imputed processes as fixed TDC's; and (iii) a joint analysis. According to the LPML statistics presented in Table 1, the CO model performs the worst in this data analysis, regardless of the method used to incorporate the longitudinal predictor (e.g. raw versus modeled) or whether parametric versus MFPT S_0 was assumed. For the two types of raw analysis, the flexibility obtained from an MFPT generalization of the log-logistic model improves predictive performance, though not dramatically so. Moreover, it is also clear that imputation and joint methods predict almost identically but are inferior to simple raw analysis in this setting. Observe from Table 2 that point estimates of β under the PO model are similar across types of analysis and that they are different for the CO model.

Tseng et al. (2005) rejected the Cox model based on a test involving Schoenfeld residuals and proposed the CO model as a plausible alternative. However, the LPML criterion suggests the Cox model has increased predictive ability over the CO model, at least in the Bayesian framework considered here. Furthermore, slightly improved prediction under the PO model was seen (explored further in Sect. 4.2), and in Sect. 4.3 we tie this to a Cox model with time-weighted trajectories that satisfies the proportional hazards assumption. From Table 1, the general conclusions about predictive model comparison drawn for this data set are that a raw LOCF analysis is preferred over imputed or joint methods, the PO model is preferred over the CO and Cox models,

Table 1 LPML across models (larger is better)

Survival model	Longitudinal trajectory	PO	Cox	CO
Log-logistic	Raw (LOCF)	−867	−870	−937
Polya tree	Raw (LOCF)	−865	−866	−938
Polya tree	Imputed	−947	−959	−973
Log-logistic	Mixed model	−947	−959	−973
Polya tree	Mixed model	−945	−956	−973

Table 2 Posterior estimates across models

Method	PO	Cox	CO
Parametric/raw	−0.75 (−1.02,−0.53)	−0.65 (−0.74,−0.56)	−0.36 (−0.44,−0.27)
MFPT/raw	−0.74 (−0.85,−0.64)	−0.64 (−0.73,−0.55)	−0.37 (−0.45,−0.29)
MFPT/imputed	−0.74 (−0.97,−0.52)	−0.37 (−0.52,−0.24)	0.16 (−0.01,0.30)
Parametric/joint	−0.78 (−1.02,−0.53)	−0.39 (−0.54,−0.25)	0.19 (0.01,0.33)
MFPT/joint	−0.79 (−1.00,−0.52)	−0.40 (−0.54,−0.24)	0.19 (0.01,0.32)

and the CO model might be excluded from further consideration, given the assumptions that were made going into the analysis.

It is worth noting that these data might not provide an ideal setting to advocate for joint analysis. First, with egg counts of this type, it might not be expected that there is much error in observation. Hence, a fundamental reason for performing a joint analysis is absent. Moreover, the observed processes might be more predictive of death since not all of the egg count trajectories fit the log gamma structure that is posited for these data. Figures 1 and 2 illustrate two distinct patterns of egg production. The fitted trajectories from PO, CO and Cox joint models are presented in these figures, along with the plot obtained from analysis of the longitudinal data by itself. The four estimates are virtually identical for both flies (they were also similar for all other flies). The log-gamma shape fits the data from fly 247 (Fig. 1) much better by any of the survival models and approaches compared to fly 22 (Fig. 2), where the simple $\{\log(t), t - 1\}$ basis oversmooths the trajectory.

Finally, we calculated predictive densities for flies 247 and 22 under all of the models and methods (Figs. 1 and 2, panels 2 and 3). Observe for fly 247 that the actual death time is highly plausible under both PO and Cox analysis with modeled trajectories, while considerably less so under the CO model; all three survival approaches fare similarly with raw LOCF. None of the approaches predict fly 22's death well, a fly that lived quite long and laid a large number of eggs relative to many other flies; see also Sect. 4.4. In general, data from flies with long lifetimes are not fit well by any of the models considered here. Insight was provided by Carey (2003, p. 63) who writes "*. . . females that exhibit high levels of lifetime reproduction must necessarily live long enough to lay eggs over a sustained period. However, this relationship becomes progressively weaker at older ages because (i) egg laying decreases at older ages thus*

reducing the rate at which lifetime totals accumulate; and (ii) inasmuch as there is no cost of reproduction, some long-lived females continue to lay eggs well into their advanced ages while others lay few or no eggs at the oldest ages. This increases the variance in lifetime reproduction of long-lived flies and therefore reduces its correlation with longevity."

4.2 Proportional odds rate

As noted in the previous section, the PO model predictively outperforms the other survival specifications, as measured by LPML, across a variety of analysis types. Also, estimates of the survival regression coefficient were similar across PO models and approaches. Both the Cox model and the Sundaram (2006) PO model have interpretations in terms of rates. Let $O_i(t) = \frac{1-S_i(t)}{S_i(t)}$ be the odds of an event at time $t > 0$ for a subject with covariate $x_i(t)$. The PO model stipulates

$$O_i'(t) = \frac{d}{dt}\, O_i(t) = e^{x_i(t)\beta} O_0'(t), \quad O_0(t) = \frac{1 - S_0(t)}{S_0(t)}.$$

Thus,

$$\frac{O_1'(t)}{O_2'(t)} = \exp[\{x_1(t) - x_2(t)\}\beta].$$

For example, if $x_1(t) = x_2(t) + \Delta$, then the time-to-event odds *are increasing* by a factor of $e^{\beta\Delta}$ for $x_1(t)$ relative to $x_2(t)$, for all t. In terms of the hazard, $O_i'(t) = h_i(t)/S_i(t)$; for either the hazard or the odds rate, larger values indicate more life is being "used up." For the medfly data, $\widehat{O}_i'(t) = [N_i(t) + 1]^{-0.74} O_0'(t)$ under the best-predicting MFPT model. A larger egg count indicates relatively less life is being used up at time t; i.e. the odds of survival increase faster with more eggs.

The hazard under the PO model is $h_i(t) = e^{x_i(t)\beta} h_0(t)/[S_0(t)\{1 + R_i(t, \beta, S_0)\}]$. The ratio of hazards is therefore

$$\frac{h_1(t)}{h_2(t)} = \exp[\{x_1(t) - x_2(t)\}\beta] \frac{1 + R_2(t, \beta, S_0)}{1 + R_1(t, \beta, S_0)},$$

where, for $i = 1, 2$, $R_i(t, \beta, S_0) = \int_0^t e^{x_i(s)\beta} O_0'(s) ds$. Consider the situation where an individual permanently changes a covariate trajectory by a constant amount Δ at time t'. For example an individual might get an organ transplant, change medication, start running, or take up recreational chainsaw juggling. We can consider two hypothetical individuals, one who did not make the change, with covariate $x_1(t)$, and one who did, with covariate

$$x_2(t) = \begin{cases} x_1(t) & \text{for } t < t' \\ x_1(t) + \Delta & \text{for } t \geq t' \end{cases}.$$

For these covariate processes, the hazard ratio jumps from unity to $h_1(t')/h_2(t') = e^{-\Delta\beta}$ at $t = t'$ but decays back to unity over time: $\lim_{t\to\infty} h_1(t)/h_2(t) = 1$. The PO model has an immediate change in hazard in the same way that the Cox model does, but this change is mediated towards one as time progresses. That is, in terms of risk, the PO model eventually 'forgets' the behavioral change whereas the Cox model does not.

As noted in Sect. 4.1, many estimated trajectories from a joint analysis with a gamma kernel random effects model are oversmoothed for the medfly data. A more flexible longitudinal model that represents a compromise between this approach and using the empirical egg counts (LOCF) might provide better predictive ability. Therefore, we consider penalized and unpenalized B-spline longitudinal models with PO for survival since it has the largest LPML among the three competing survival models. Specifically, $K = 20$ quadratic B-spline basis functions $\{\psi_1(\cdot), \ldots, \psi_K(\cdot)\}$ on 19 knots equispaced from zero to 100 days, roughly the span of lifetimes in the medfly data set, define each trajectory through coefficients $\mathbf{b}_i = (b_{i1}, \ldots, b_{iK})$. Each egg-laying trajectory has a B-spline representation, namely $x_i(t) = \sum_{k=1}^{20} b_{ik}\psi_k(t)$ with $\mathbf{b}_i|\boldsymbol{\mu}, \boldsymbol{\Sigma} \sim N_{20}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and prior $\boldsymbol{\mu} \sim N_{20}(\mathbf{0}, 1000\mathbf{I})$ independent of $\boldsymbol{\Sigma}^{-1} \sim \text{Wish}(20, 0.05\mathbf{I})$ so that $E(\boldsymbol{\Sigma}^{-1}) = \mathbf{I}$. Most log-egg count trajectories vary between 0 and 5, and so assuming $\text{var}(b_{ij}) \approx 1$ *a priori* is rather generous in terms of allowing for sudden jumps in basis function coefficients. This specification yields LPML $= -879$ for the parametric joint model, a bit worse than LOCF but much better than using the basis $\{\log(t), t - 1\}$.

An interesting phenomenon was observed in the PO, B-spline setting with priors that supported very large coefficients: the LPML values exceeded those obtained from using raw trajectories. With $\boldsymbol{\Sigma}^{-1} \sim \text{Wish}(20, 0.5(10^{-j})\mathbf{I})$ for $j = 1, 2, 3, 4, 5, 6$, the following LPML values were obtained: -879, -868, -849, -834, -828 and -826. An examination of the estimated trajectories indicates that this prior sequence increasingly results in modeled profiles that dipped markedly below the last observed count, and often below zero (the lower bound for the trajectory) at the event time. This has the effect of increasing the density near the observed event time and hence increasing the CPO for an individual fly, thereby increasing the overall aggregate LPML. For the two-parameter gamma kernel model, the survival portion of the model modifies trajectories negligibly (Figs. 1–6). However, for this set of rich profile classes, there are many parameters for which there is little or no egg laying data providing information, and therefore the spline coefficients near the event time can be disproportionately affected by the survival portion of the model. That is, only a few days worth of egg-laying information is going into the spline parameter for those basis functions near the event time; see Appendix B for how this relates to posterior sampling.

To tame this anomalous behavior of the egg-laying trajectory near death, we further considered a penalized B-spline expansion for each trajectory, with all medflies sharing the same penalty, or smoothness term λ; i.e. λ is estimated from the whole of the data and so all trajectories more or less have the same level of "smoothness." Specifically, $x_i(t) = \mu_i + \sum_{k=1}^{20} b_{ik}\psi_k(t)$ where $p(\mathbf{b}_i|\lambda) \propto \lambda^{(K-1)/2}\exp(-0.5\lambda\mathbf{b}_i'\boldsymbol{\Omega}\mathbf{b}_i)$. The matrix $\boldsymbol{\Omega}$ obtains from a first order difference penalty, see Lang and Brezger (2004) and Appendix B. The prior $\lambda \sim \Gamma(a_\lambda, b_\lambda)$ was assumed with $a_\lambda = b_\lambda = 0.1$. The posterior median of λ is 0.19 and a 95% posterior interval (0.18–0.20) is highly

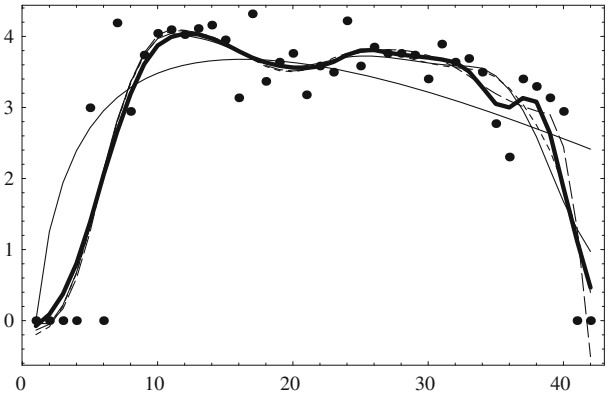

Fig. 7 Fly 19 (event time 42 days). Raw egg laying counts (*dots*), and log-gamma kernel basis (*oversmoothed, thin solid*). Unpenalized splines are *solid*, *short*, and *long-dashed*; penalized spline is *thick solid*

concentrated relative to the prior, for both the parametric and MFPT analysis. The resulting LPML statistics are -869 for MFPT and -872 for parametric, only slightly worse than using the raw trajectories, and an improvement over the unpenalized spline with the (reasonable) prior specified above. Increasing to $K = 40$ gives LPML values of -868 for the MFPT and -874 for the parametric model. It would seem that the added penalized B-spline flexibility does not provide better predictive performance over simply using LOCF, but it comes close.

Consider the data for fly 19 in Fig. 7. Plotted are the estimated trajectories from the parametric PO model using the log-gamma kernel basis, the unpenalized spline basis with parameters $20^{-1}\mathbf{I}$, $2000^{-1}\mathbf{I}$, and $200000^{-1}\mathbf{I}$, and penalized B-spline with $K = 40$. These yield CPO_{19} values of -3.2 (log-gamma kernel), -2.3 (raw), -2.2 (penalized), and -2.5, -2.1, -1.9 for the unpenalized splines. The CPO statistics are ordered according to where the trajectories hit the vertical line at $t = 42$, the death time for this fly. Clearly, trajectory behavior near the event time can greatly impact prediction accuracy of the event time, and thus a "correct" model for trajectories in the joint setting is extremely important in developing a predictive model for these data.

4.3 Cox model

Although proportional hazards is rejected for the raw trajectories according to the test proposed by Therneau and Grambsch (2000), the Cox model better predicts survival for these data than the CO model. Given β and S_0, the Cox model specifies $h(t) = (N(t) + 1)^{\beta} h_0(t)$. The negative regression estimate ($\hat{\beta} = -0.65$; Table 2) indicates decreased hazard of death with increasing egg production. The baseline hazard for the Cox model is essentially zero until about 20 days, so the decreased hazard is practically interpretable only beyond 20 days.

The Cox model is the 'default' choice for many joint modeling applications, being used in almost all papers devoted to the subject. Therefore, it is of interest to see if a

modified version of the Cox model can predictively outperform the PO model. To this end, we considered various 'fixes' in an attempt to improve prediction and remedy the apparent rejection of proportional hazards.

Almost half the trajectories (120) have one or more zeros right before death. Perhaps simply having a function indicating zero eggs laid, for instance $I\{y_i(t) > 0\}$, might better predict mortality. Or perhaps it's the size of a drop or increase, $y_i(t) - y_i(t-1)$, from one day to the next, or a lagged count, $y_i(t-L)$, that better predicts death times. We looked at the following transformed trajectories: (a) a first difference, $y_i^*(t) = y_i(t) - y_i(t-1)$, which approximates the derivative $y_i'(t)$ gives LPML $= -963$; (b) the trajectory lagged by one day, $y_i^*(t) = y_i(t-1)$, gives LPML $= -908$; (c) a simple indicator of whether the fly produces eggs, $y_i^*(t) = I\{y_i(t) > 0\}$, gives LPML $= -974$; (d) the integrated trajectory $y_i^*(t) = \int_0^t y_i(s)ds$ gives LPML $= -975$, worse than not including covariates at all; and (e), the integrated trajectory standardized to have volume one, $y_i^*(t) = [\int_0^t y_i(s)ds]/[\int_0^{T_i} y_i(s)ds]$, gives LPML $= -971$.

None of these approaches provide better prediction than the untransformed version; most have LPML values close to -974, the value obtained by removing covariates altogether. However, a plot of the scaled Schoenfeld residuals roughly indicates a larger initial (negative) effect due to $y_i(t) = \log\{N_i(t) + 1\}$ attenuating to zero somewhat linearly by 100 days. Correspondingly, we also fit the weighted trajectory $y_i^*(t) = [1 - t/100]y_i(t)$. With this weighted transformation, the proportional hazards assumption is not violated according to the test suggested by Therneau and Grambsch (2000) ($p = 0.474$ with $g(t) = t$ on p. 131), and improves prediction slightly beyond PO: LPML $= -864$ for both parametric and MFPT. Note that the weighted trajectory yields the same qualitative result as the PO model: the ratio of any two hazard functions approaches one as $t \to 100$, giving further evidence that fecundity is more predictive early on.

A natural approach here would be to consider a Cox model with time-varying regression effects in the raw setting (Martinussen and Scheike, 2006, Sect. 6.3). A joint modeling specification allowing time-varying coefficients was considered by Song and Wang (2008). For completeness, we fit a varying-coefficient model assuming LOCF

$$h_i(t) = e^{\beta(t)y_i(t)}h_0(t), \text{ where } \beta(t) = \sum_{k=1}^{20} c_i\kappa^{-1}\phi\left\{-0.5(l_k - t)^2/\kappa^2\right\}.$$

We assumed $\mathbf{c} = (c_1, \ldots, c_{20})' \sim N_{20}(\mathbf{0}, \mathbf{I}\,10^2)$ and $\log(\kappa) \sim N(3.0, 0.25)$, yielding LPML=$-865$ for both the parametric and MFPT models. This prior on κ reflects a belief in the smoothness of $\beta(t)$ that the standard deviation of the Gaussian kernel ranges from 7 to 55 days with 95% probability. This implies a correlation $\text{corr}\{\beta(t), \beta(t+\Delta)\}$ of approximately 0.9, 0.6, 0.4, 0.3 for $\Delta = 10$, 20, 30, and 40 days. The number of local extrema of $\beta(t)$ over $t \in (0, 100)$ ranges from one to about six, and a 99% prior probability interval for $\beta(t)$ includes $(-3, 3)$ at the edges $t = 0$ and $t = 100$ days, and $(-4, 4)$ in the middle. Different priors, including a hyperprior on $\text{var}(c_i)$, yielded LPML values in the range -865 to -876. The vector $(\mathbf{c}', \theta_1, \theta_2)$ was updated in one large M-H step; κ was updated in a separate M-H step. The knots $l_k = (k-1)100/19$

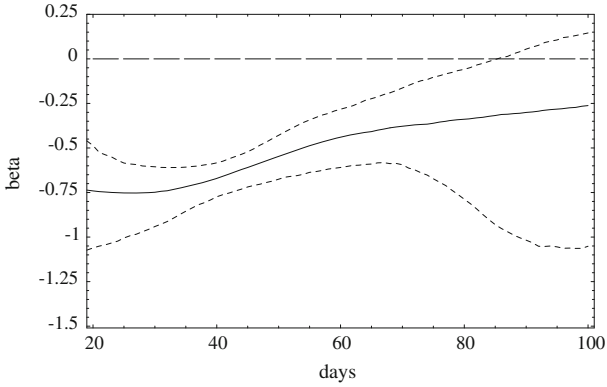

Fig. 8 Posterior mean $\hat{\beta}(t)$ and 95% CI versus days for the MFPT Cox model

were equispaced, ranging from $l_1 = 0$ to $l_{20} = 100$. The time-varying coefficient model predicts about as well as using time-weighted trajectories dying to zero by time 100. Correspondingly, Fig. 8 shows roughly a linear increase in $\hat{\beta}(t)$ over time, with much greater variability near 100 days. Only 10 flies lived past 70 days, 4 of these lived past 80 days.

Müller et al. (2001) find for a different subset of flies ($n = 531$) that it is not the height of the trajectory, but rather its shape that is associated with survival. They considered imputed trajectories

$$y_i(t) = I\{t \geq b_{i1}\} b_{i3} \exp\{-b_{i2}(t - b_{i1})\}$$

in the Cox model. This function is zero until the changepoint b_{i1}, at which point it decays exponentially from height b_{i3} with rate b_{i2}. The changepoint b_{i1} was taken to be the time at which egg-laying was a maximum, with (b_{i2}, b_{i3}) estimated via least-squares. The predictor was standardized to have maximal height one, and these authors found this changepoint model, which measures "reproductive exhaustion," to fit better than other predictors the authors considered, under the AIC criterion. Along these lines, we looked at normalized raw trajectories, $y_i^*(t) = y_i(t)/\max\{y_i(t) : t = 1, \ldots, T_i\}$, obtaining LPML $= -867$ from the parametric model. The normalized trajectory improves prediction, confirming findings from Müller et al. (2001). Note that such trajectories may be useful for explaining survival, but cannot be used for prediction unless one has prior knowledge of the maximum.

A related model introduced by De Blasi and Hjort (2007) posits proportional hazards but with a logistic link:

$$h(t) = \frac{e^{x(t)\beta}}{(1 + e^{x(t)\beta})^{\kappa}} h_0(t).$$

The traditional Cox model with log-link obtains for $\kappa = 0$, and a logistic link, implying bounded risk, obtains when $\kappa = 1$. We fit this model with $\kappa = 1$ using raw TDC's

in a parametric log-logistic survival model and obtained LPML $= -870$, the same as the Cox model with an exponential link; the MFPT model gave LPML $= -867$.

4.4 Cox and Oakes AFT model

The log-logistic model with no predictors has LPML $= -974$. This indicates that for our analysis, the CO model with imputed or jointly modeled TDC's has about the same predictive capacity as not including the covariates at all. We examine specific flies with the greatest ratio of CPO values to infer the types of trajectories that are best supported under each model.

Fly 38 (Fig. 3) and fly 225 both support the jointly modeled CO model over the raw trajectories, and these flies have similar trajectories and density estimates. In both cases there are a large number (>15) of zero counts before initial egg production starts. Although this causes the simple gamma kernel basis to "shoot off into space" in estimating the longitudinal profile, the joint model predicts these survival times relatively well. It may be that these types of trajectories are the driving force for a positive estimate of the joint model regression coefficient. In contrast, the death time of fly 45 (Fig. 4) is better predicted by the empirical egg counts. Here, only a few days of no egg production occur before egg-laying starts, and there is one day with no eggs just prior to death. This one day, not captured by the gamma kernel basis, causes a dramatic spike in the predictive density for the raw analysis right before death, therefore increasing the predictive ability of the raw model.

Flies 22 (Fig. 2) and 111 (Fig. 6) support the Cox over the CO model when raw trajectories are used. Both of these flies initially have a large number of days with no eggs, followed by egg-laying activity, but they have at least one zero count right before death. The large number of initial zeroes causes the CO model to "eat up" a large amount of survival time relative to baseline, leaving little probability mass past 20 days. Fly 6 (Fig. 5) and fly 66 support the CO model over Cox under LOCF. Both have an initial flurry of egg-laying in the first 25 days, followed by 10–20 days of no egg production, followed by one or two days of egg production right before death. For both models most probability mass has been eaten up near the time of death, but less so for the CO analysis; thus more of a probability 'spike' caused by these few zero counts is allowed under the CO model. Neither model predicts either event time well, and both have similarly shaped predictive densities.

Flies 38 (Fig. 3) and 111 (Fig. 6) are better predicted by the PO versus the CO model when using raw trajectories. Fly 38, not fit well by either model, has an increasing estimated trajectory over the lifespan. The spike at the event time is due to fixing $y_{38}(t) = 0$ for $t \geq 58$, the event time; i.e. the spike occurs after death due to assuming $y_{38}(t) = 0$. Both the PO and Cox density estimates are similar for fly 111, with a spike corresponding to one day with no eggs produced right before death. Flies 6 and 146 support the CO model over the PO model. The story is identical to that for the Cox analysis above.

Finally, comparing PO to Cox; fly 14 supports PO the most, but is ill-fit by both models with an event time of 65. Fly 23 supports Cox over PO, and has the largest event time in the data set at 99 days. No model reasonably supports flies with the longest lifetimes.

Tseng et al. (2005) conducted a frequentist joint CO analysis of the medfly data and obtained a statistically significant negative regression coefficient with $\hat{\beta} = -0.434$. In contrast, we obtained a significantly positive effect ($\hat{\beta} = 0.19$) from the joint model. A possible source for this difference appears to lie in the longitudinal data used by Tseng et al., who added two additional zero counts at the end of each fly's egg-laying profile and shifted the remaining counts two days into the past (as seen in their Fig. 3, representing flies 6, 7, 15, 10 in the original data, clockwise from upper left; flies 5, 6, 9, and 7 in reduced data set of size $n = 251$). These two additional zeroes cause a relatively steeper dip in each trajectory near the observed survival time, and allow the modeled trajectories to more closely mirror the raw data. When we shift the egg-laying trajectory one day into the future we obtain LPML $= -949$ and $\hat{\beta} = -0.66$ for the jointly modeled parametric approach, and LPML $= -696$ with $\hat{\beta} = -1.76$ for the parametric approach with raw covariates. Shifting by two days drops the LPML and decreases $\hat{\beta}$ further in both cases.

Clearly, from a predictive point of view, the longitudinal basis is very important. The nonlinear bases chosen for use with these data allow for negative trajectory values, which is not possible for the response $\log\{N_i(t) + 1\}$. The basis $\{\log(t), t - 1\}$ also allows for trajectories that continue to increase for all time. When we fit the joint CO model with the simple linear basis $\{1, t\}$, the regression coefficient β was significantly positive with LPML $= -900$ for the MFPT and parametric models, considerably better than the nonlinear basis.

5 Discussion

A major deficiency to date in the joint modeling literature as we see it is the lack of comparison among semiparametric models. Our analysis of the medfly data illustrates the potential utility of considering parametric and alternative semiparametric families of models in addition to the Cox model, and different modes of incorporating TDC's in survival analysis.

The Bayesian modeling process taken here involves the use of a single nonparametric prior for the baseline survival distribution in CO, PO, and Cox models, which enabled, with relative ease, a means of comparing these approaches in terms of prediction accuracy. The particular choice of an MFPT prior has a number of nice features, including the embedding of a standard parametric family of survival models within the nonparametric class. This Bayesian approach is quite flexible, both in terms of the ability to center a nonparametric prior that is used in a variety of survival models on the same parametric family, and allowing for a robust longitudinal model with a relatively large dimension for $\mathbf{b}_i$. In this paper the dimension of $\mathbf{b}_i$ ranged from two to forty. A flexible non Bayesian approach to survival analysis was recently developed by Zhang and Davidian (2008), who introduced a family of distributions based on a class of polynomials and fit Cox, PO, and AFT models using this 'quasi-parametric' approach allowing for ease of model comparison. In their supplementary online material, they also fit the CO model with TDC's.

For the most fertile medflies, the PO model provided better prediction than the Cox model, but marginally so. The CO model was ruled out under all scenarios. When

the effect of fecundity on survival was allowed to taper off in time, either through varying coefficients or by using time-weighted raw coefficients, the Cox model provided prediction about as good as or a bit better than PO. Regardless, the effect of egg-laying on survival decreases as the fly ages. Of additional interest was the impact of jointly modeling the egg-laying trajectories, either through the basis $\{\log(t), t-1\}$ used in Tseng et al. (2005) or richer bases. For these data, the raw trajectories using LOCF provided the best prediction, although splines did almost as well. Surprising to us was the rather large impact the unpenalized spline basis had on LPML when using relatively flat priors. Highly flexible bases should be carefully used for longitudinal modeling in the joint setting as the survival portion of the model could have an unrealistic impact in the absence of some sort of penalty or smoothness constraint.

Our analysis of the medfly data makes perhaps the obvious point that joint modeling may not always be necessary, even if the longitudinal process under consideration is decidedly stochastic. Rice (2004) conducted an exploratory analysis of the medfly data using nonparametric smoothing techniques for functional and longitudinal data. One method he used to assess the predictiveness of egg production on lifespan was to smooth the first 30 days of egg laying and subsequently predict lifetime using the mean of nearest neighbor lifetimes. This approach to forecasting medfly death times was deemed by Rice to have limited predictive accuracy. Müller and Stadtmüller (2005) considered a similar task only their goal was to use egg-laying productivity in early life (first 30 days) to predict whether a fly will have a short or long lifespan (binary response). They also smoothed the predictor process, which was then used in a nonparametric functional binomial regression model. Similar to Rice (2004), the predictive capacity of their approach was found to be relatively low (cross-validation prediction errors between 35 and 48%), due in large part to an overall lack of transparent differences in predictor processes for short and long lived flies. However, their analysis concluded that the single most important predictor of longevity in medflies was high egg productivity in later life. In studies like these, more research to determine when it is desirable to do joint modeling versus a raw or two-stage analysis, perhaps based on smoothed trajectories, would be a welcome addition to the literature. In particular, incorporating functional data analysis methods that can naturally identify many aspects of trajectory shape may improve prediction of survival times.

In terms of predictive performance, for this particular data set, within a modeling paradigm (LOCF, imputed, modeled, and MFPT versus parametric) the choice of PO, Cox, or CO changed the LPML measure drastically. The survival model represents gross, overarching assumptions about the data-generating mechanism. Using either jointly modeled trajectories, or imputed values based on the longitudinal model greatly reduced the predictive utility of the models. The MFPT generalization is akin to adding detail to an initially washed canvas, for these data the MFPT generalization apparently adds little in the way of prediction.

Acknowledgements The authors thank Professor Yi-Kuan Tseng for providing $\widehat{\mathbf{b}}_i$ under the model fit in Tseng et al. (2005) and mapping them to the original data.

References

Aalen OO (1980) A model for nonparametric regression analysis of counting processes. In: Klonecki W, Kozek A, Rosinski J (eds) Mathematical statistics and probability theory. Lecture notes in statistics. Springer, New York, pp 1–25

Banerjee S, Gelfand AE, Finley AO, Sang H (2008) Gaussian predictive process models for large spatial datasets. J Roy Stat Soc Ser B 70:825–848

Bedrick EJ, Christensen R, Johnson WO (2000) Bayesian accelerated failure time analysis with application to veterinary epidemiology. Stat Med 19:221–237

Berger JO, Guglielmi A (2001) Bayesian testing of a parametric model versus nonparametric alternatives. J Am Stat Assoc 96:174–184

Branscum AJ, Hanson TE (2008) Bayesian nonparametric meta-analysis using Polya tree mixture models. Biometrics 64:825–833

Brown ER, Ibrahim JG (2003) A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics 59:221–228

Brown ER, Ibrahim JG, DeGruttola V (2005) A flexible B-spline model for multiple longitudinal biomarkers and survival. Biometrics 61:64–73

Bycott P, Taylor J (1998) A comparison of smoothing techniques for CD4 data measured with error in a time-dependent Cox proportional hazards model. Stat Med 17:2061–2077

Carey JR (2003) Longevity: the biology and demography of life span. Princeton University Press, Princeton

Carey JR, Liedo P, Müller HG, Wang JL, Chiou JM (1998) Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of Mediterranean fruit fly females. J Gerontol A Biol Sci Med Sci 53:245–251

Chen M-H, Shao Q-M, Xu D (2002) Sufficient and necessary conditions on the propriety of posterior distributions for generalized linear mixed models. Sankhya Ser A 64:57–85

Chen M-H, Ibrahim JG, Shao Q-M (2006) Posterior propriety and computation for the Cox regression model with applications to missing covariates. Biometrika 93:791–807

Chiou JM, Müller HG, Wang JL, Carey JR (2003) A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. Stat Sinica 13:1119–1133

Cox DR (1972) Regression models and life-tables (with discussion). J Roy Stat Soc Ser B 34:187–220

Cox DR, Oakes D (1984) Analysis of survival data. Chapman and Hall, London

De Blasi P, Hjort NL (2007) Bayesian survival analysis in proportional hazard models with logistic relative risk. Scand J Stat 34:229–257

Fahrmeir L, Kneib T (2009) Propriety of posteriors in structured additive regression models: theory and empirical evidence. J Stat Plann Infer 139:843–859

Faucett CL, Thomas DC (1996) Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. Stat Med 15:1663–1685

Ferguson TS (1974) Prior distributions on spaces of probability measures. Ann Stat 2:615–629

Geisser S, Eddy WF (1979) A predictive approach to model selection. J Am Stat Assoc 74:153–160

Hanson TE (2006) Inference for mixtures of finite Polya tree models. J Am Stat Assoc 101:1548–1565

Hanson T, Johnson WO (2002) Modeling regression error with a mixture of Polya trees. J Am Stat Assoc 97:1020–1033

Hanson TE, Yang M (2007) Bayesian semiparametric proportional odds models. Biometrics 63:88–95

Hanson T, Johnson WO, Laud P (2009) A unified approach to semiparametric inference for survival models with step-stress covariates. Can J Stat 37:60–79

Hsieh F, Tseng Y-K, Wang J-L (2006) Joint modeling of survival and longitudinal data: Likelihood approach revisited. Biometrics 62:1037–1043

Higdon D (2001) Space and space-time modeling using process convolutions. Discussion paper 2001-2003, Institute for Statistics and Decision Sciences, Duke University.

Ibrahim JG, Chen M-H, Sinha D (2001) Bayesian survival analysis. Springer-Verlag, New York

Kneib T (2006) Mixed model based inference in structured additive regression. PhD Thesis, Munich University

Laird NM, Ware JH (1982) Random-effects models for longitudinal data. Biometrics 38:963–974
Lang S, Brezger A (2004) Bayesian P-splines. J Comput Graph Stat 13:183–212
Lavine M (1992) Some aspects of Polya tree distributions for statistical modeling. Ann Stat 20:1222–1235
Lavine M (1994) More aspects of Polya tree distributions for statistical modeling. Ann Stat 22:1161–1176
Law NJ, Taylor JMG, Sandler H (2002) The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure. Biostatistics 3:547–563
Li Y, Lin X, Müller P (2009) Bayesian inference in semiparametric mixed models for longitudinal data. Biometrics. doi:10.1111/j.1541-0420.2009.01227.x
Martinussen T, Scheike TH (2006) Dynamic regression models for survival analysis. Springer, New York
Mauldin RD, Sudderth WD, Williams SC (1992) Polya trees and random distributions. Ann Stat 20:1203–1221
Müller H-G, Stadtmüller U (2005) Generalized functional linear models. Ann Stat 33:774–805
Müller H-G, Carey JR, Wu D, Liedo P, Vaupel JW (2001) Reproductive potential predicts longevity of female Mediterranean fruit flies. Proc Roy Soc Lond B 268:445–450
Paddock SM, Ruggeri F, Lavine M, West M (2003) Randomized Polya tree models for nonparametric Bayesian inference. Stat Sinica 13:443–460
Prentice RL (1982) Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika 69:331–342
Prentice RL, Kalbfleisch JD (1979) Hazard rate models with covariates. Biometrics 35:25–39
Rice JA (2004) Functional and longitudinal data analysis: perspectives on smoothing. Stat Sinica 14:631–647
Schoop R, Graf E, Schumacher M (2008) Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. Biometrics 64:603–610
Song X, Wang CY (2008) Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. Biometrics 64:557–566
Song X, Davidian M, Tsiatis AA (2002) An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. Biostatistics 3:511–528
Sundaram S (2006) Semiparametric inference in proportional odds model with time-dependent covariates. J Stat Plann Infer 136:320–334
Therneau TM, Grambsch PM (2000) Modeling survival data: extending the Cox model. Springer, New York
Tseng Y-K, Hsieh F, Wang J-L (2005) Joint modelling of accelerated failure time and longitudinal data. Biometrika 92:587–603
Tsiatis AA, Davidian M (2004) Joint modeling of longitudinal and time-to-event data: an overview. Stat Sinica 14:809–834
Wang Y, Taylor JMG (2001) Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. J Am Stat Assoc 96:895–905
Wulfsohn MS, Tsiatis AA (1997) A joint model for survival and longitudinal data measured with error. Biometrics 53:330–339
Yang R, Chen M-H (1995) Bayesian analysis for random coefficient regression models using noninformative priors. J Multivar Anal 55:283–311
Zhang M, Davidian M (2008) "Smooth" semiparametric regression analysis for arbitrarily censored time-to-event data. Biometrics 64:567–576