A general framework for parametric survival analysis

Michael J. Crowther[1][*][†] **and Paul C. Lambert**[1,2]

Parametric survival models are being increasingly used as an alternative to the Cox model in biomedical research. Through direct modelling of the baseline hazard function we can gain greater understanding of the risk profile of patients over time, obtaining absolute measures of risk. Commonly used parametric survival models, such as the Weibull, make restrictive assumptions of the baseline hazard function, such as monotonicity, which is often violated in clinical datasets. In this article, we extend the general framework of parametric survival models proposed by Crowther and Lambert (2013), to incorporate relative survival, and robust and cluster robust standard errors. We describe the general framework through three applications to clinical datasets, in particular, illustrating the use of restricted cubic splines, modelled on the log hazard scale, to provide a highly flexible survival modelling framework. Through the use of restricted cubic splines we can derive the cumulative hazard function analytically beyond the boundary knots, resulting in a combined analytic/numerical approach, which substantially improves the estimation process compared to only using numerical integration. User friendly Stata software is provided which significantly extends parametric survival models available in standard software. Copyright © 0000 John Wiley & Sons, Ltd.

Keywords: survival analysis, parametric modelling, Gaussian quadrature, maximum likelihood, splines, time-dependent effects, relative survival

1. Introduction

The use of parametric survival models is growing in applied research [1, 2, 3, 4, 5], as the benefits become recognised and the availability of more flexible methods becomes available in standard software. Through a parametric approach, we can obtain clinically useful measures of absolute risk allowing greater understanding of individual patient risk profiles [6, 7, 8], particularly important with the growing interest in personalised medicine. A model of the baseline hazard or survival allows us to calculate absolute risk predictions over time, for example in prognostic models, and enables the translation of hazards ratios back to the absolute scale, for example when calculating the number needed to treat. In addition, parametric models are especially useful for modelling time-dependent effects [9, 4], and when extrapolating survival [10, 11].

Commonly used parametric survival models, such as the exponential, Weibull and Gompertz proportional hazards models, make strong assumptions about the shape of the baseline hazard function. For example, the Weibull model assumes a monotonically increasing or decreasing baseline hazard. Such assumptions restrict the underlying function

[1] *University of Leicester, Department of Health Sciences, Adrian Building, University Road, Leicester LE1 7RH.*
[2] *Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, S-171 77 Stockholm, Sweden.*
[*] *Correspondence to: Michael J. Crowther. Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH.*
[†] *E-mail: michael.crowther@le.ac.uk*

that can be captured, and are often simply not flexible enough to capture those observed in clinical datasets which often exhibit turning points in the underlying hazard function [12, 13].

Crowther and Lambert [14] recently described the implementation of a general framework for the parametric analysis of survival data, which allowed any well-defined hazard or log hazard function to be specified, with the model estimated using maximum likelihood utilising Gaussian quadrature. In this article we extend the framework to relative survival, and also allow for robust and cluster robust standard errors. In particular, throughout this article we concentrate on the use of restricted cubic splines to demonstrate the framework, and describe a combined analytic/numeric approach to greatly improve the estimation process.

Various types of splines have been used in the analysis of survival data, predominantly on the hazard scale which results in an analytically tractable cumulative hazard function. For example, M-splines, which by definition are non-negative can be directly applied on the hazard scale, due to the positivity condition. Kooperberg et al. (1995) proposed using various types of splines on the log hazard scale, such as piecewise linear splines [15, 16]. In this article we use restricted cubic splines to model the log hazard function, which by definition ensures the hazard function is positive across follow-up, but has the computational disadvantage that the cumulative hazard requires numerical integration to calculate it. Restricted cubic splines have been used widely within the flexible parametric survival modelling framework of Royston and Parmar [17, 18], which are modelled on the log cumulative hazard scale. The switch to the log cumulative hazard scale provides analytically tractable cumulative hazard and hazard functions; however, when there are multiple time-dependent effects there are difficulties in interpretation of time-dependent hazard ratios, since these will vary over different covariate patterns, even with no interaction between these covariates [18].

In Section 2, we derive the general framework and extend it to incorporate cluster robust standard errors and incorporate background mortality for the extension to relative survival. In Section 3, we describe a special case of the framework using restricted cubic splines to model the baseline hazard and time-dependent effects, and describe how the estimation process can be improved through a combined analytical and numerical approach. In Section 4 we apply the spline based hazard models to datasets in breast and bladder cancer, illustrating the improved estimation routine, the application of relative survival, and the use of cluster robust standard errors, respectively. We conclude the paper in Section 5 with a discussion.

2. A general framework for the parametric analysis of survival data

We begin with some notation. For the ith patient, where $i = 1, \ldots, N$, we define t_i to be the observed survival time, where $t_i = \min(t_i^*, c_i)$, the minimum of the true survival time, t_i^*, and the censoring time, c_i. We define an event indicator d_i, which takes the value of 1 if $t_i^* \leq c_i$ and 0 otherwise. Finally, we define t_{0i} to be the entry time for the ith patient, i.e. the time at which a patient becomes at risk.

Under a parametric survival model, subject to right censoring and possible delayed entry (left truncation); the overall log-likelihood function can be written as follows

$$l = \sum_{i=1}^{N} \log L_i \tag{1}$$

with log likelihood contribution for the ith patient

$$\log L_i = \log \left\{ f(t_i)^{d_i} \left(\frac{S(t_i)}{S(t_{0i})} \right)^{1-d_i} \right\}$$

$$= d_i \log\{f(t_i)\} + (1 - d_i) \log\{S(t_i)\} - (1 - d_i) \log\{S(t_{0i})\} \tag{2}$$

where $f(t_i)$ is the probability density function and $S(.)$ is the survival function. If $t_{0i} = 0$, the third term of Equation (2) can be dropped. Using the relationship

$$f(t) = h(t) \times S(t) \tag{3}$$

where $h(t)$ is the hazard function at time t, substituting Equation (3) into (2), we can write

$$\log L_i = \log \left\{ h(t_i)^{d_i} \frac{S(t_i)}{S(t_{0i})} \right\}$$

$$= d_i \log\{h(t_i)\} + \log\{S(t_i)\} - \log\{S(t_{0i})\} \tag{4}$$

Now given that

$$S(t) = \exp\left(-\int_0^t h(u)\mathrm{d}u\right) \tag{5}$$

we can write Equation (4) entirely in terms of the hazard function, $h(.)$, incorporating delayed entry

$$\log L_i = d_i \log\{h(t_i)\} - \int_{t_{0i}}^{t_i} h(u)\mathrm{d}u \tag{6}$$

The log-likelihood formulation of Equation (6) implies that, if we specify a well-defined hazard function, where $h(t) > 0$ for $t > 0$, and can subsequently integrate it to obtain the cumulative hazard function, we can then maximise the likelihood and fit our parametric survival model using standard techniques [19].

When a standard parametric distribution is chosen, for example the exponential, Weibull or Gompertz, and for the moment assuming proportional hazards; we can directly integrate the hazard function to obtain a closed form expression for the cumulative hazard function. As described in Section 1, these distributions are simply not flexible enough to capture many observed hazard functions. If we postulate a more flexible function for the baseline hazard, which cannot be directly integrated analytically, or wish to incorporate complex time-dependent effects for example; we then require numerical integration techniques in order to maximise the likelihood.

2.1. Numerical integration using Gaussian quadrature

Gaussian quadrature is a method of numerical integration which provides an approximation to an analytically intractable integral [20]. It turns an integral into a weighted summation of a function evaluated at a set of pre-defined points called quadrature nodes or abscissae. Consider the integral from Equation (6)

$$\int_{t_{0i}}^{t_i} h(u)\mathrm{d}u \tag{7}$$

To obtain an approximation of the integral through Gaussian quadrature, we first must undertake a change of interval using

$$\int_{t_{0i}}^{t_i} h(u)\mathrm{d}u = \frac{t_i - t_{0i}}{2}\int_{-1}^{1} h\left(\frac{t_i - t_{0i}}{2}z + \frac{t_{0i} + t_i}{2}\right)\mathrm{d}z \tag{8}$$

Applying numerical quadrature, in this case Gauss-Legendre, results in

$$\int_{t_{0i}}^{t_i} h(u)\mathrm{d}u \approx \frac{t_i - t_{0i}}{2}\sum_{j=1}^{m} v_j h\left(\frac{t_i - t_{0i}}{2}z_j + \frac{t_{0i} + t_i}{2}\right) \tag{9}$$

where $\boldsymbol{v} = \{v_1, \ldots, v_m\}$ and $\boldsymbol{z} = \{z_1, \ldots, z_m\}$ are sets of weights and node locations, respectively, with m the number of quadrature nodes. Under Gauss-Legendre quadrature the weights $v_j = 1$. We must specify the number of quadrature nodes, m, with the numerical accuracy of the approximation dependent on m. As with all methods which use numerical integration, the accuracy of the approximation can be assessed by comparing estimates with an increasing number of nodes. We return to the issue of choosing the number of quadrature points in Section 3.

2.2. Excess mortality models

In population-based studies where interest lies in mortality associated with a particular disease, it is not always possible to use cause of death information. This may be due to this information not being available or it considered too unreliable to use [21, 22]. In these situations it is common to model and estimate excess mortality by comparing the mortality experienced amongst a diseased population to that expected amongst a disease free population. The methods have most commonly been applied to population-based cancer studies, but have also been used in studies of HIV [23] and cardiovascular disease [24]. The total mortality (hazard) rate, $h_i(t)$, is partitioned into the expected mortality rate, $h_i^*(t)$, and the excess mortality rate associated with a diagnosis of disease, $\lambda_i(t)$.

$$h_i(t) = h_i^*(t) + \lambda_i(t) \tag{10}$$

The expected mortality rate, $h_i^*(t)$, is usually obtained from national or regional life tables stratified by age, calendar year, sex and sometimes other covariates such as socio-economic class [25].

Transforming to the survival scale gives,

$$S_i(t) = S_i^*(t)R_i(t) \tag{11}$$

where $R_i(t)$ is known as the relative survival function and $S_i^*(t)$ is the expected survival function.

The effect of covariates on the excess mortality rate is usually considered to be multiplicative and so covariates, $\boldsymbol{X}_i$ are modelled as,

$$h_i(t) = h_i^*(t) + \lambda_0(t) \exp(\boldsymbol{X}_i \boldsymbol{\beta}) \tag{12}$$

where λ_0 is the baseline excess hazard function and $\boldsymbol{\beta}$ is a vector of log excess hazard ratios (also referred to as log excess mortality rate ratios). This model assumes proportional excess hazards, but in population-based cancer studies this assumption is rarely true and there has been substantial work on methods to fit models that relax the assumption of proportionality [24, 26, 27, 28].

A common model for analysing excess mortality is an extension of Royston-Parmar models [24]. These models are fitted on the log cumulative excess hazard scale. With multiple time-dependent effects interpretation of hazard ratios can be complicated and so there are advantages to modelling on the log hazard scale instead. For example, in a model on the log cumulative excess hazard scale where both age group and sex are modelled as time-dependent effects, but with no interaction between the covariates, the estimated hazard ratio for sex would be different in each of the age groups. In a model on the log excess hazard scale, this would not be the case [18]. Previous work by Remontet et al. (2007) used numerical integration, but used quadratic splines, limited to only two knots, with no restriction on the splines [29].

The log-likelihood for an excess mortality model is as follows,

$$\log L_i = d_i \log \{h^*(t_i) + \lambda(t_i)\} + \log \{S^*(t_i)\} + \log \{R(t_i)\} - \log \{S^*(t_{0i})\} - \log \{R(t_{0i})\} \tag{13}$$

Since the terms $\log \{S^*(t_i)\}$ and $\log \{S^*(t_{0i})\}$ do not depend on any model parameters they can be omitted from the likelihood function for purposes of estimation. This means that in order to estimate the model parameters the expected mortality rate at the time of death, $h^*(t_i)$, is needed for subjects that experience an event.

2.3. Cluster robust standard errors

In standard survival analysis we generally make the assumption that observations are independent; however, in some circumstances we can expect observations to be correlated if a group structure exists within the data. For example in the analysis of recurrent event data, where individual patients can experience an event multiple times, resulting in multiple observations per individual. In this circumstance, we would expect observations to be correlated within groups. Failing to account for this sort of structure can underestimate standard errors.

Given $\hat{\boldsymbol{V}}$, our standard estimate of the variance covariance matrix, which is the inverse of the negative Hessian matrix evaluated at the maximum likelihood estimates, we define the robust variance estimate developed by Huber (1987) and White (1980,1982) [30, 31, 32]

$$V_r = \hat{\boldsymbol{V}} \left(\sum_{i=1}^{N} \boldsymbol{u}_i' \boldsymbol{u}_i \right) \hat{\boldsymbol{V}} \tag{14}$$

where $\boldsymbol{u}_i$ is the contribution of the ith observation to $\partial \log L / \partial \beta$, with N the total number of observations.

This can be extended to allow for a clustered structure. Suppose the N observations can be classified into M groups, which we denote by $G_1, \ldots, G_M$, where groups are now assumed independent rather than individual level observations. The robust estimate of variance becomes

$$V_r = \hat{\boldsymbol{V}} \left(\sum_{j=1}^{M} \boldsymbol{u}_j^{(G)'} \boldsymbol{u}_j^{(G)} \right) \hat{\boldsymbol{V}} \tag{15}$$

where $\boldsymbol{u}_i^{(G)}$ is the contribution of the jth group to $\partial \log \mathrm{L} / \partial \beta$. More specifically, Rogers (1993) [33] noted that if the log-likelihood is additive at the observation level, where

$$\log L = \sum_{i}^{N} \log L_i \tag{16}$$

then with $\boldsymbol{u}_i = \partial \log L_i / \partial \beta$, we have

$$\boldsymbol{u}_j^G = \sum_{i \in G_j} \boldsymbol{u}_i \tag{17}$$

We follow the implementation in Stata which also incorporates a finite sample adjustment of $\hat{V}_r^* = \{M/(M-1)\}\hat{V}_r$.

3. Improving the estimation when using restricted cubic splines

The very nature of the modelling framework described above implies that we can specify practically any general function in the definition of our hazard or log hazard function, given that it satisfies $h(t) > 0$ for all $t > 0$. To illustrate the framework, we concentrate on a particular flexible way of modelling survival data, using restricted cubic splines [34].

We begin by assuming a proportional hazards model, modelling the baseline log hazard function using restricted cubic splines

$$\log h_i(t) = \log h_0(t) + \boldsymbol{X}_i\boldsymbol{\beta} = s(\log(t)|\boldsymbol{\gamma}, \boldsymbol{k}_0) + \boldsymbol{X}_i\boldsymbol{\beta} \tag{18}$$

where $\boldsymbol{X}_i$ is a vector of baseline covariates with associated log hazard ratios $\boldsymbol{\beta}$, and $s(\log(t)|\boldsymbol{\gamma}, \boldsymbol{k}_0)$ is a function of $\log(t)$ expanded into restricted cubic spline basis with knot location vector, $\boldsymbol{k}_0$, and associated coefficient vector, $\boldsymbol{\gamma}$. For example, if we let $u = \log(t)$, and with knot vector, $\boldsymbol{k}_0$

$$s(u|\boldsymbol{\gamma}, \boldsymbol{k}_0) = \gamma_0 + \gamma_1 s_1 + \gamma_2 s_2 + \cdots + \gamma_{m+1} s_{m+1} \tag{19}$$

with parameter vector $\boldsymbol{\gamma}$, and derived variables s_j (known as the basis functions), where

$$s_1 = u \tag{20}$$

$$s_j = (u - k_j)_+^3 - \lambda_j(u - k_{\min})_+^3 - (1 - \lambda_j)(u - k_{\max})_+^3 \tag{21}$$

where for $j = 2, \ldots, m + 1$, $(u - k_j)_+^3$ is equal to $(u - k_j)^3$ if the value is positive and 0 otherwise, and

$$\lambda_j = \frac{k_{\max} - k_j}{k_{\max} - k_{\min}} \tag{22}$$

In terms of knot locations, for the internal knots, we use by default the centiles of the uncensored log survival times, and for the boundary knots we use the minimum and maximum observed uncensored log survival times. The restricted nature of the function imposes the constraint that the fitted function is linear beyond the boundary knots, ensuring a sensible functional form in the tails where often data is sparse. The choice of the number of spline terms (more spline terms allows greater flexibility), is left to the user. A recent extensive simulation study assessed the use of model selection criteria to select the optimum degrees of freedom within the Royston-Parmar model (restricted cubic splines on the log cumulative hazard scale), which showed no bias in terms of hazard ratios, hazard rates and survival functions, with a reasonable number of knots as guided by AIC/BIC [13].

3.1. Complex time-dependent effects

Time-dependent effects, i.e. non-proportional hazards, are commonplace in the analysis of survival data, where covariate effects can vary over prolonged follow-up time, for example in the analysis of registry data [9]. Continuing with the special case of using restricted cubic splines, we can incorporate time-dependent effects into our model framework as follows

$$\log h_i(t) = s(\log(t)|\boldsymbol{\gamma}_0, \boldsymbol{k}_0) + \boldsymbol{X}_i\boldsymbol{\beta} + \sum_{p=1}^{P} x_{ip} s(\log(t)|\boldsymbol{\gamma}_p, \boldsymbol{k}_p) \tag{23}$$

where for the pth time-dependent effect, with $p = \{1, \ldots, P\}$, we have x_p, the pth covariate, multiplied by some spline function of log time, $s(\log(t)|\boldsymbol{\gamma}_p, \boldsymbol{k}_p)$, with knot location vector, $\boldsymbol{k}_p$, and coefficient vector, $\boldsymbol{\gamma}_p$. Once again, degrees of freedom, i.e. number of knots, for each time-dependent effect can be guided using model selection criteria, and/or the impact of different knot locations assessed through sensitivity analysis.

3.2. Improving estimation

Given that the modelling framework is extremely general, in that the numerical integration can be applied to a wide range of user-defined hazard functions, the application of Gaussian quadrature to estimate the models may not be the most computationally efficient. For example, in Crowther and Lambert [14], we compared a Weibull proportional hazards model, with the equivalent general hazard model using numerical integration.

In the restricted cubic spline based models described above, the restricted nature of the spline function forces the baseline log hazard function to be linear beyond the boundary knots. In those areas the cumulative hazard function can actually be written analytically, as the log hazard is a linear function of log time. Defining our boundary knots to be k_{01}, k_{0n}, we need only conduct numerical integration between k_{01}, k_{0n}, using the analytical form of the cumulative hazard function beyond the boundary knots.

We define δ_{0i} and δ_{1i} to be the intercept and slope of the log hazard function for the ith patient before the first knot, k_{01}, and ϕ_{0i} and ϕ_{1i} to be the intercept and slope of the log hazard function for the ith patient after the final knot, k_{0n}. If there are no time-dependent effects then $\{\delta_{0i}, \delta_{1i}, \phi_{0i}, \phi_{1i}\}$ are constant across patients. The cumulative hazard function can then be defined in three components

$$H_i(t) = H_{1i}(t) + H_{2i}(t) + H_{3i}(t) \tag{24}$$

If we assume $t_{0i} < k_{01}$ and $t_i > k_{0n}$, then before the first knot, we have

$$H_{1i}(t) = \frac{\exp(\delta_{0i})}{\delta_{1i} + 1} \left\{ \min(t_i, k_{01})^{\delta_{1i}+1} - t_{0i}^{\delta_{1i}+1} \right\} \tag{25}$$

and after the final knot, we have

$$H_{3i}(t) = \frac{\exp(\phi_{0i})}{\phi_{1i} + 1} \left\{ t_i^{\phi_{1i}+1} - \max(t_{0i}, k_{0n})^{\phi_{1i}+1} \right\} \tag{26}$$

and $H_{2i}(t)$ becomes

$$H_{2i}(t) \approx \frac{k_{0n} - k_{01}}{2} \sum_{j=1}^{m} v_j h_i \left(\frac{k_{0n} - k_{01}}{2} z_j + \frac{k_{01} + k_{0n}}{2} \right) \tag{27}$$

The alternative forms of the cumulative hazard function for situations where, for example, $t_{0i} > k_{01}$ are detailed in Appendix A. This combined analytical/numerical approach allows us to use far fewer quadrature nodes, which given numerical integration techniques are generally computationally intensive, is a desirable aspect of the estimation routine. We illustrate this in Section 4.1.

3.3. Improving efficiency

In this section we conduct a small simulation study to compare the efficiency of the Kaplan-Meier estimate of the survival function, to a parametric formulation using splines, in particular, when data is sparse in the right tail. We simulate survival times from a Weibull distribution with scale and shape values of 0.2 and 1.3, respectively. Censoring times are generated from a U(0,6) distribution, with the observed survival time taken as the minimum of the censoring and event times, and an administrative censoring time of 5 years. This provides a realistic combination of intermittent and administrative censoring. A thousand repetitions are conducted, each with a sample size of 200.

In each repetition, we calculate the Kaplan-Meier estimate, and associated standard error, of survival at 4 and 5 years, and the parametric equivalent using a spline based model with 3 degrees of freedom. The median number of events across the simulations was 101, with a median of 5 events during the final year of follow-up. Results are presented in Table 1.

Table 1. Bias and mean squared error of $\log(-\log(S(t)))$ at 4 and 5 years.

Time		Kaplan-Meier	Parametric model
4 years	Bias	-0.0019	-0.0038
	MSE	0.1251	0.1100
5 years	Bias	0.0066	0.0063
	MSE	0.1565	0.1481

From Table 1, we see that at both 4 and 5 years, the mean squared error is lower for the parametric approach, compared to the Kaplan-Meier estimate. Bias is essentially negligible for all estimates. This indicates a gain in efficiency for the parametric approach in this particular scenario. Of course, this simulation setting is limited to a simple case of a Weibull, but note that we do not fit the correct parametric model, but an incorrect flexible model still does better than the Kaplan-Meier.

4. Example applications

We aim to show the versatility of the framework through three different survival modelling areas, utilising splines, whilst providing example code in the appendix to demonstrate the ease of implementation to researchers.

4.1. Breast cancer survival

We begin with a dataset of 9721 women aged under 50 and diagnosed with breast cancer in England and Wales between 1986 and 1990. Our event of interest is death from any cause, where 2,847 events were observed, and we have restricted follow-up to 5 years, leading to 6,850 censored at 5 years. We are interested in the effect of deprivation status, which was categorised into 5 levels; however, in this example we restrict our analyses to comparing the least and most deprived groups. We subsequently have a binary covariate, with 0 for the least deprived and 1 for the most deprived group.

In this section we wish to establish the benefit of incorporating the analytic components, described in Section 3.2, compared to the general method of only using numerical integration, described in Section 2. We use the general Stata software package, `stgenreg`, described previously [14], to fit the full quadrature based approach, and a newly developed Stata package, `strcs`, which implements the combined analytic and numeric approach when using splines on the log hazard scale. We apply the spline based models shown in Equation (18), with 5 degrees of freedom (6 knots), i.e. 5 spline variables to capture the baseline, incorporating the proportional effect of deprivation status, with an increasing number of quadrature points, until estimates are found to have converged to 3, 4 and finally 5 decimal places.

Table 2. Comparison of estimates when using different numbers of nodes for the fully numeric approach

Parameter	Number of Nodes								
	10	20	30	40	50	100	250	500	1000
deprivation	0.268560	0.269302	0.269363	0.269380	0.269386	0.269393	0.269395	0.269395	0.269395
	(0.039203)	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)
γ_0	-2.916819	-2.912434	-2.910463	-2.909648	-2.909240	-2.908601	-2.908289	-2.908201	-2.908162
	(0.060860)	(0.060749)	(0.060701)	(0.060682)	(0.060673)	(0.060659)	(0.060651)	(0.060648)	(0.060647)
γ_1	-0.085113	-0.066088	-0.062178	-0.060704	-0.059979	-0.058850	-0.058346	-0.058214	-0.058158
	(0.027644)	(0.027508)	(0.027460)	(0.027442)	(0.027432)	(0.027416)	(0.027408)	(0.027405)	(0.027404)
γ_2	0.038085	0.072033	0.078483	0.080923	0.082146	0.084099	0.084980	0.085214	0.085314
	(0.019940)	(0.019462)	(0.019297)	(0.019231)	(0.019196)	(0.019135)	(0.019101)	(0.019090)	(0.019084)
γ_3	0.147381	0.121891	0.115869	0.113473	0.112252	0.110276	0.109344	0.109088	0.108976
	(0.018258)	(0.017899)	(0.017675)	(0.017569)	(0.017509)	(0.017398)	(0.017333)	(0.017311)	(0.017299)
γ_4	-0.040437	-0.027974	-0.025152	-0.024017	-0.023433	-0.022474	-0.022017	-0.021890	-0.021834
	(0.014469)	(0.014429)	(0.014372)	(0.014343)	(0.014327)	(0.014296)	(0.014277)	(0.014270)	(0.014267)
γ_5	0.010185	0.003174	0.001279	0.000518	0.000133	-0.000481	-0.000775	-0.000857	-0.000893
	(0.013512)	(0.013438)	(0.013408)	(0.013395)	(0.013388)	(0.013374)	(0.013366)	(0.013363)	(0.013361)
log-likelihood	-8739.9490	-8753.8333	-8756.2213	-8757.0858	-8757.5006	-8758.1249	-8758.3830	-8758.4444	-8758.4683

Standard errors in parentheses

Table 3. Comparison of estimates when using different numbers of nodes for the combined analytical/numeric approach

| Parameter | Number of Nodes | | | | | | | | | |
	10	20	30	40	50	100	250	500	1000
deprivation	0.269295	0.269376	0.269390	0.269393	0.269394	0.269395	0.269395	0.269395	0.269395
	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)	(0.039202)
γ_0	-2.906390	-2.908770	-2.908353	-2.908198	-2.908148	-2.908133	-2.908133	-2.908133	-2.908133
	(0.060656)	(0.060663)	(0.060650)	(0.060648)	(0.060647)	(0.060647)	(0.060647)	(0.060647)	(0.060647)
γ_1	-0.061499	-0.059304	-0.058469	-0.058225	-0.058149	-0.058118	-0.058117	-0.058117	-0.058117
	(0.027397)	(0.027411)	(0.027405)	(0.027404)	(0.027404)	(0.027403)	(0.027403)	(0.027403)	(0.027403)
γ_2	0.077581	0.083720	0.084902	0.085233	0.085337	0.085390	0.085390	0.085390	0.085390
	(0.019033)	(0.019082)	(0.019082)	(0.019080)	(0.019080)	(0.019079)	(0.019079)	(0.019079)	(0.019079)
γ_3	0.112949	0.110410	0.109370	0.109043	0.108938	0.108889	0.108888	0.108888	0.108888
	(0.017117)	(0.017279)	(0.017291)	(0.017290)	(0.017289)	(0.017288)	(0.017288)	(0.017288)	(0.017288)
γ_4	-0.024649	-0.022456	-0.021996	-0.021857	-0.021812	-0.021790	-0.021790	-0.021790	-0.021790
	(0.014188)	(0.014258)	(0.014263)	(0.014263)	(0.014263)	(0.014263)	(0.014263)	(0.014263)	(0.014263)
γ_5	-0.000164	-0.000367	-0.000745	-0.000869	-0.000908	-0.000921	-0.000922	-0.000922	-0.000922
	(0.013428)	(0.013363)	(0.013360)	(0.013360)	(0.013360)	(0.013360)	(0.013360)	(0.013360)	(0.013360)
log-likelihood	-8754.2660	-8757.6342	-8758.2559	-8758.4167	-8758.4634	-8758.4839	-8758.4840	-8758.4840	-8758.4840

Standard errors in parentheses

Table 2 compares parameter estimates and standard errors under the full numerical approach, across varying number of quadrature nodes, and Table 3 presents the equivalent results for the combined analytic/numeric approach. From Table 2, we still observe variation in estimates and the log likelihood to 5 or 6 decimal places between 500 and 1000 nodes, whilst for the combined approach shown in Table 3, the maximum difference between 100 and 1000 nodes is 0.000001. For the combined approach the log-likelihood does not change to 3 decimal places between 100 and 1000 nodes, whilst the log-likelihood for the full numerical approach is only the same to 1 decimal place.

We found that with the full numerical approach it required 23 nodes, and 50 nodes, to establish consistent estimates to 3 and 4 decimal places, respectively. We compare that to 18 nodes, and 27 nodes under the combined analytic and numerical approach. Final results for the combined approach using 27 nodes are presented in Table 4.

Table 4. Results from combined analytic/numerical spline based survival model.

Variable	Hazard ratio	95% CI	
Deprivation (most)	1.309	1.212	1.414
Baseline	Coefficient	95% CI	
γ_1	-0.059	-0.112	-0.005
γ_2	0.085	0.047	0.122
γ_3	0.110	0.076	0.143
γ_4	-0.022	-0.050	0.006
γ_5	-0.001	-0.027	0.025
Intercept	-2.908	-3.027	-2.789

From Table 4 we observe a statistically significant hazard ratio of 1.309 (95% CI: 1.212, 1.414), indicating an increased hazard rate in the most deprived group, compared to the least deprived. Comparing computation time, the general approach with 49 quadrature nodes took 20.5 seconds on a standard laptop, compared to 17.5 using the combined approach with 27 nodes.

Figure 1 shows the fitted survival functions from the full numerical approach (using `stgenreg`), the combined analytic/numerical approach (using `strcs`), and the Cox model. It is clear that all three models yield essentially identical fitted survival functions, though from a visual inspection all three appear to fit poorly.
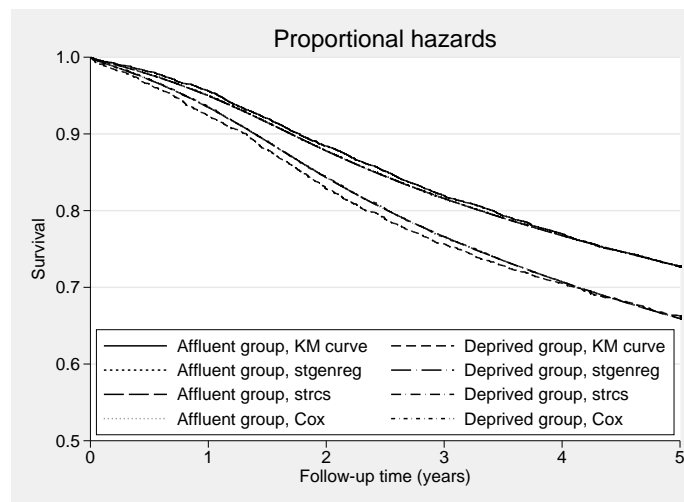

Figure 1. Kaplan-Meier curve by deprivation status, with fitted survival functions overlaid, from `stgenreg`, `strcs` and Cox models.

We can investigate the presence of a time-dependent effect due to deprivation status, by applying Equation (23). We use 5 degrees of freedom to capture the baseline and use 3 degrees of freedom to model the time-dependent effect of deprivation status. Figure 2 shows the time-dependent hazard ratio, illustrating the decrease in the effect of deprivation over time.

The improved fit when incorporating the time-dependent effect of deprivation status is illustrated in Figure 3.
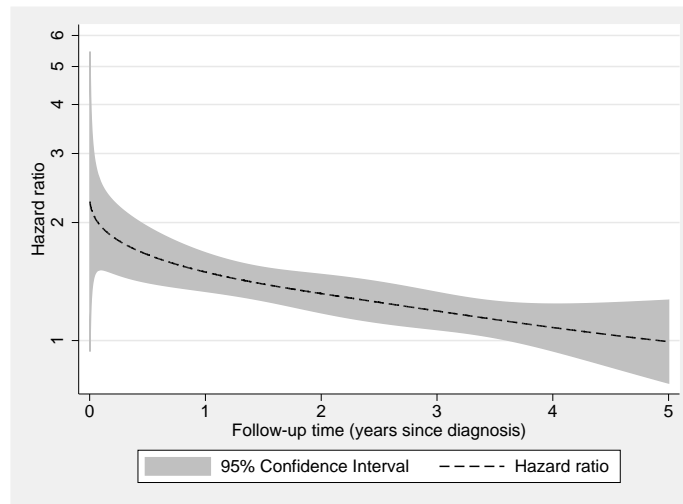
Figure 2. Time-dependent hazard ratio for deprivation status.
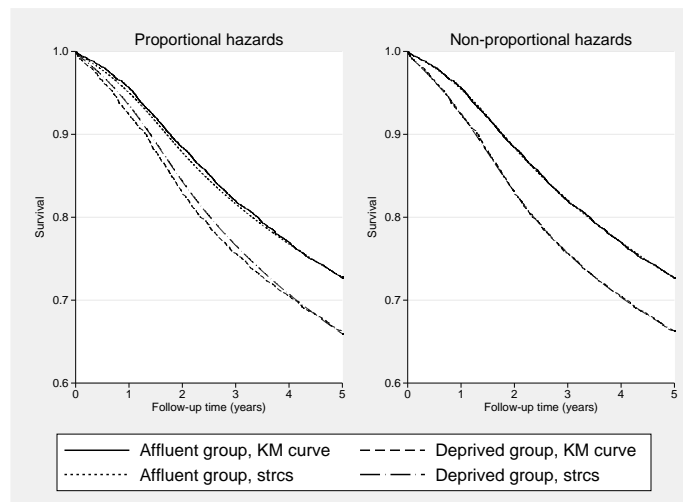

Figure 3. Fitted survival function overlaid on the Kaplan-Meier curve, under proportional hazards and non-proportional hazards models using `strcs`.

Example Stata code to fit time-independent and time-dependent models presented in this section are included in Appendix B.

4.2. Excess mortality model

For the excess mortality model we use the same data source as in Section 4.1. However, we now include women aged over 50. Expected mortality is stratified by age, sex, calendar year, region and deprivation quintile [25]. As for the analysis in Section 4.1, we only include the least and most deprived groups for simplicity. Age is categorised into 5 groups, < 50, 50-59, 60-69, 70-79, 80+. There are 41,645 subjects included in the analysis, with a total of 17,057 events before 5 years post diagnosis.

4.2.1. Proportional excess hazards model We initially fit a model where the excess mortality rate is assumed to be proportional between different covariate patterns. We compare the estimates to a model using restricted cubic splines on the log cumulative hazard scale [24]. In both models 6 knots are used with these placed evenly according to the distribution of log death times.

From Table 5, we observe very similar hazard ratios and their 95% confidence intervals between the models on different scales.

Table 5. Comparison of excess hazard ratios (and 95% confidence intervals) from models with the linear predictor on the log hazard scale and the log cumulative hazard scale. Both models have 6 knots with these placed evenly according to the distribution of log death times.

Covariate	log hazard	log cumulative hazard
Deprivation (most)	1.313	1.313
	(1.265,1.364)	(1.265,1.364)
Age (50-59)	1.055	1.055
	(0.998,1.114)	(0.998,1.114)
Age (60-69)	1.071	1.071
	(1.014,1.130)	(1.015,1.131)
Age (70-79)	1.453	1.454
	(1.372,1.539)	(1.373,1.540)
Age (80+)	2.647	2.647
	(2.484,2.822)	(2.484,2.821)

Age (< 50) is the reference group

4.2.2. Time-dependent effects A model is now fitted where the assumption of proportional excess hazards is relaxed for all covariates. This is done by incorporating an interaction between each covariate and a restricted cubic spline function of log time with 4 knots (3 degrees of freedom). The knots are placed evenly according to the distribution of log death times. The estimated excess hazard ratio for deprivation group can be seen in Figure 4. As there is not an interaction between deprivation group and age group then this hazard ratio is assumed to apply for each of the 5 age groups. If the model was fitted on the log cumulative excess hazard scale, then this would not be the case. This is illustrated in Figure 5 where the same linear predictor has been fitted for a model on the log cumulative excess hazard scale and the estimated excess hazard ratio is shown for two age groups and is shown to be different.
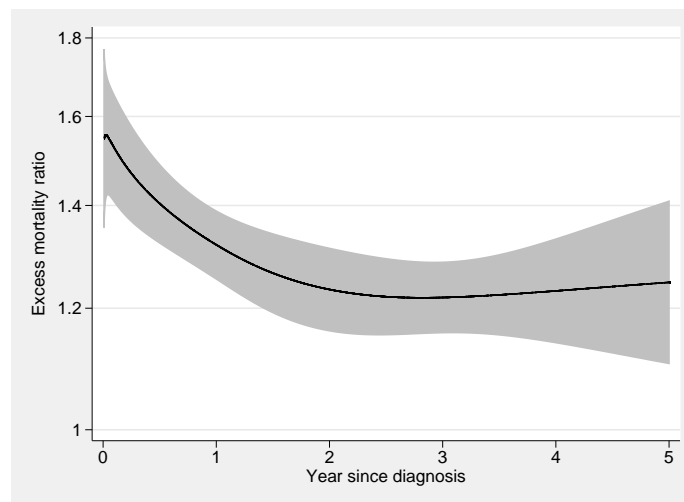

Figure 4. Excess hazard ratio comparing most deprived with least deprived group. The model used 6 knots for the baseline and 4 knots for the time-dependent effect.

The impact of the default interior knot locations can be assessed through sensitivity analyses, varying the knot locations. In Figure 6, we compare the default choice (interior knots at 1.024 and 2.660), to three other choices, illustrating some minor variation in the tails of the estimated shape of the time-dependent excess hazard ratio; however, the functional form is generally quite robust to knot location.

Example Stata code to fit time-independent and time-dependent excess mortality models presented in this section are included in Appendix C.
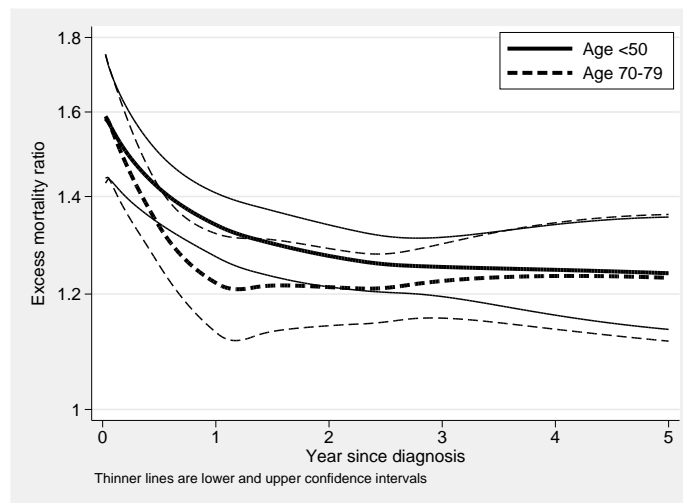
Figure 5. Excess hazard ratios comparing most deprived with least deprived group. The model used 6 knots for the baseline and 4 knots for the time-dependent effect.
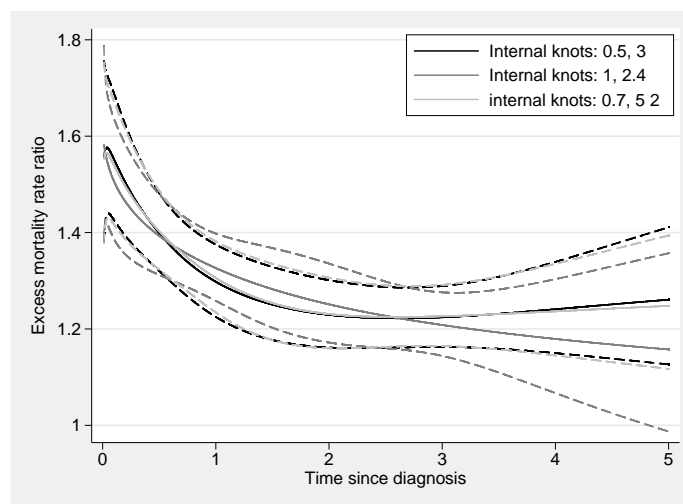

Figure 6. Excess hazard ratios comparing most deprived with least deprived group. The model used 6 knots for the baseline and 4 knots for the time-dependent effect, with 3 choices for the interior knots of the time-dependent effect. Dashed lines indicate 95% confidence intervals.

4.3. Cluster robust errors

To illustrate the use of cluster robust standard errors, we use a dataset of 85 patients with bladder cancer [35, 36]. We fit a model for recurrent event data, where the event of interest is recurrence of bladder cancer. Each patient can experience a total of 4 events, shown in Table 6. A total of 112 events were observed. Covariates of interest include treatment group (0 for placebo, 1 for thiotepa), initial number of tumors (range 1 to 8, with 8 meaning 8 or more), and initial size of tumors (in centimetres, with range 1 to 7).

Table 6. Number of patients who were censored or experienced up to 4 recurrences of bladder cancer

Recurrence number	Number of patients		
	Censored	Event	Total
1	38	47	85
2	17	29	46
3	5	22	27
4	6	14	20

To allow for the inherent structure, events nested within patients, we fit a parametric version of the Prentice-Williams-Peterson model, allowing for cluster robust standard errors. This model uses non-overlapping time intervals, thus for example, a patient is not at risk of a second recurrence until after the first has occurred. The baseline hazard for each event is allowed to vary, i.e. there is a stratification factor by event. We use 5 knots for a shared baseline between the events, but allow departures from this baseline using restricted cubic splines with 3 knots for each of the subsequent events. For comparison, we also fit a Cox model, stratified by event number, with cluster robust standard errors [37]. Results are presented in Table 7.

Table 7. Results from spline based and Cox models with cluster robust standard errors.

Variable	Spline hazard model				Cox model			
	Hazard Ratio	Robust Std. Err.	95% CI		Hazard Ratio	Robust Std. Err.	95% CI	
group	0.699	0.149	0.459	1.063	0.716	0.148	0.478	1.073
size	0.990	0.064	0.872	1.123	0.992	0.061	0.878	1.120
number	1.146	0.060	1.035	1.269	1.127	0.058	1.018	1.247

From Table 7, we observe similar estimates from the spline based model, compared to the Cox model with cluster robust standard errors. We can compare estimated baseline hazard rates for each of the four ordered events, from the spline based model, shown in Figure 7.
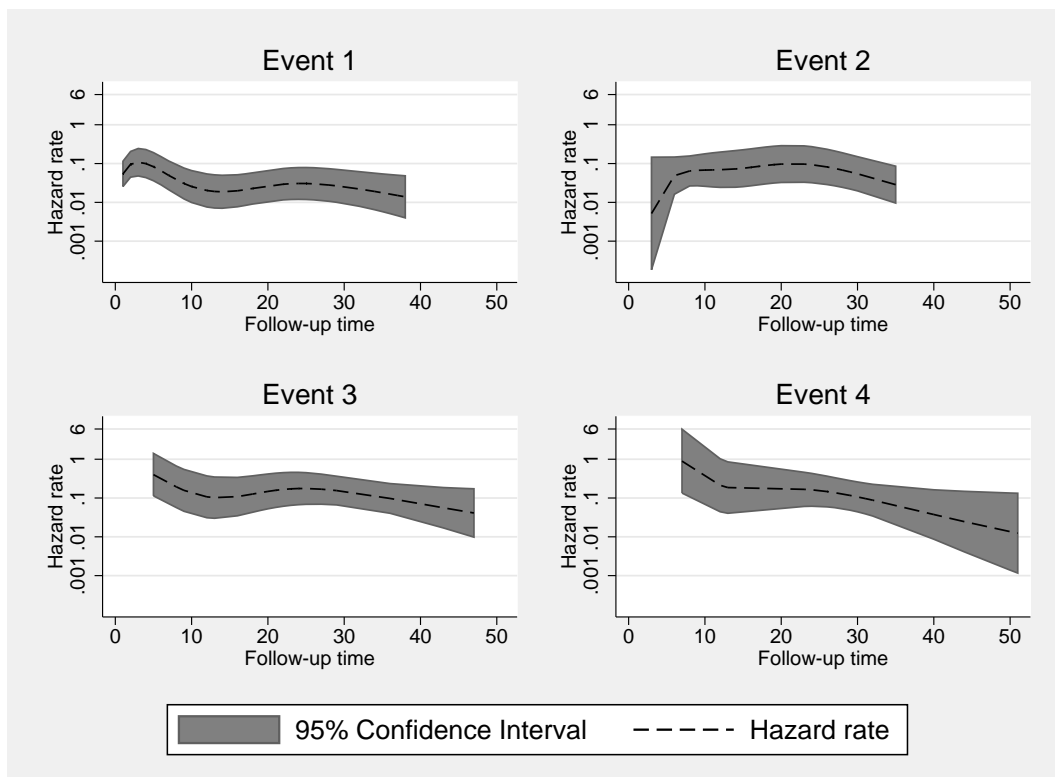

Figure 7. Baseline hazard rates for the four ordered events.

We can see from Figure 7, that those patients who go on to experience a third and fourth event have a high initial hazard rate, demonstrating the fact that they will likely be a more severe subgroup.

Example Stata code to fit the cluster robust spline model is included in Appendix D.

5. Discussion

We have described a general framework for the parametric analysis of survival data, incorporating any combination of complex baseline hazard functions, time-dependent effects, time-varying covariates, delayed entry (left truncation), robust and cluster robust standard errors, and the extension to relative survival. Modelling the

baseline hazard, and any time-dependent effects parametrically, can offer a greater insight into the risk profile over time. Parametric modelling is of particular importance when extrapolating survival data, for example within an economic decision modelling framework [11]. In this article we concentrated on the use of restricted cubic splines, which offer great flexibility to capture the observed data, but also a likely sensible extrapolation if required, given the linear restriction beyond the boundary knots.

In particular, we described how the general framework can be optimised in special cases with respect to the estimation routine, utilising the restricted nature of the splines to incorporate the analytic parts of the cumulative hazard function, in combination with the numerical integration. This provided a much more efficient estimation process, requiring far fewer quadrature nodes to obtain consistent estimates, providing computational benefits. However, it is important to note that although we have concentrated on the use of splines in this article, essentially any parametric function can be used to model the baseline (log) hazard function and time-dependent effects.

In application to the breast cancer data, we showed that the general numerical approach requires a large number of quadrature nodes, compared to the combined analytic/numeric approach, in order to obtain consistent estimates. This is due to the numerical approach struggling to capture high hazards at the beginning of follow-up time. Given that hazard ratios are usually only reported to two/three decimal places, the large number of nodes used in Section 4.1 will often not be required. In further examples not shown, where the hazard is low at the beginning of follow-up, often < 30 nodes are sufficient with the full numerical approach.

We have chosen to use restricted cubic spline functions of log time, since in many applications we have found this to provide an equivalent or better fit, compared to using splines of time. However, in studies with age as the timescale it may be more appropriate to use spline functions of untransformed time.

Other approaches to modelling the baseline hazard and time-dependent effects include using the piecewise exponential framework, either through a Bayesian [38] or classical approach [39]. Han et al. (2014) [39] developed a reduced piecewise exponential approach which can be used to identify shifts in the hazard rate over time based on an exact likelihood ratio test, a backward elimination procedure, and an optional presumed order restriction on the hazard rate; however, can be considered more of a descriptive tool, as covariates cannot currently be included. The piecewise approach assumes the baseline and any time-dependent effects follow a step-function. Alternatively, using splines, as described in this article, would produce a more plausible estimated function in continuous time, with particular benefits in terms of prediction both in and out of sample, compared to the piecewise approach.

In this article we have only looked at fixed effect survival models; however, future work involves the incorporation of frailty distributions. User friendly Stata software, written by the authors, is provided which significantly extends the range of available methods for the parametric analysis of survival data [14].

Appendix

Appendix A

For the ith patient, we have entry and survival times, t_{0i} and t_i, respectively. We define δ_{0i} and δ_{1i} to be the intercept and slope of the log hazard function for the ith patient before the first knot, k_{01}, and ϕ_{0i} and ϕ_{1i} to be the intercept and slope of the log hazard function for the ith patient after the final knot, k_{0n}. The cumulative hazard function can then be defined in three components

$$H_i(t) = H_{1i}(t) + H_{2i}(t) + H_{3i}(t)$$

If we assume $t_{0i} < k_{01}$ and $t_i > k_{0n}$, then before the first knot, we have

$$H_{1i}(t) = \frac{\exp(\delta_{0i})}{\delta_{1i} + 1} \left\{ \min(t_i, k_{01})^{\delta_{1i}+1} - t_{0i}^{\delta_{1i}+1} \right\}$$

and after the final knot, we have

$$H_{3i}(t) = \frac{\exp(\phi_{0i})}{\phi_{1i} + 1} \left\{ t_i^{\phi_{1i}+1} - \max(t_{0i}, k_{0n})^{\phi_{1i}+1} \right\}$$

and $H_{2i}(t)$ becomes

$$H_{2i}(t) \approx \frac{k_{0n} - k_{01}}{2} \sum_{j=1}^{m} v_j h_i \left(\frac{k_{0n} - k_{01}}{2} z_j + \frac{k_{01} + k_{0n}}{2} \right)$$

Alternatively, we may have observations where $k_{0n} > t_{0i} > k_{01}$ and $t_i > k_{0n}$ then we have

$$H_{1i}(t) = 0$$

$$H_{2i}(t) \approx \frac{k_{0n} - t_{0i}}{2} \sum_{j=1}^{m} v_j h_i \left(\frac{k_{0n} - t_{0i}}{2} z_j + \frac{t_{0i} + k_{0n}}{2} \right)$$

$$H_{3i}(t) = \frac{\exp(\phi_{0i})}{\phi_{1i} + 1} \left\{ t_i^{\phi_{1i}+1} - \max(t_{0i}, k_{0n})^{\phi_{1i}+1} \right\}$$

If $t_{0i} < k_{01}$ and $k_{01} < t_i < k_{0n}$, we have

$$H_{1i}(t) = \frac{\exp(\delta_{0i})}{\delta_{1i} + 1} \left\{ \min(t_i, k_{01})^{\delta_{1i}+1} - t_{0i}^{\delta_{1i}+1} \right\}$$

$$H_{2i}(t) \approx \frac{t_i - k_{01}}{2} \sum_{j=1}^{m} v_j h_i \left(\frac{t_i - k_{01}}{2} z_j + \frac{k_{01} + t_i}{2} \right)$$

$$H_{3i}(t) = 0$$

If $k_{01} < t_{0i} < t_i < k_{0n}$, we have

$$H_{1i}(t) = 0$$

$$H_{2i}(t) \approx \frac{t_i - t_{0i}}{2} \sum_{j=1}^{m} v_j h_i \left(\frac{t_i - t_{0i}}{2} z_j + \frac{t_{0i} + t_i}{2} \right)$$

$$H_{3i}(t) = 0$$

If $t_{0i} < t_i < k_{01}$, then

$$H_{1i}(t) = \frac{\exp(\delta_{0i})}{\delta_{1i} + 1} \left\{ t_i^{\delta_{1i}+1} - t_{0i}^{\delta_{1i}+1} \right\}$$

$$H_{2i}(t) = 0$$

$$H_{3i}(t) = 0$$

Finally, if $k_{0n} < t_{0i} < t_i$, we have

$$H_{1i}(t) = 0$$

$$H_{2i}(t) = 0$$

$$H_{3i}(t) = \frac{\exp(\phi_{0i})}{\phi_{1i} + 1} \left\{ t_i^{\phi_{1i}+1} - t_{0i}^{\phi_{1i}+1} \right\}$$

Appendix B

Example Stata code using 5 spline variables to model the baseline. **stgenreg** uses the full numerical approach, and **strcs** uses the combined analytic/numeric approach.

```
. stgenreg, loghazard([xb]) xb(dep5 | #rcs(df(5)) ) nodes(50)
. strcs dep5, df(5) nodes(27)
```

Incorporating a time-dependent effect:

```
. stgenreg, loghazard([xb]) xb(dep5 | #rcs(df(5)) | dep5:*#rcs(df(3))) nodes(50)
. strcs dep5, df(5) nodes(50) tvc(dep5) dftvc(3)
```

Appendix C

Example code to fit the combined analytic/numeric approach assuming proportional excess hazards, and non-proportional excess hazards

```
. strcs dep5 agegrp2 agegrp3 agegrp4 agegrp5, df(5) nodes(50) bhazard(rate)
. strcs dep5 agegrp2 agegrp3 agegrp4 agegrp5, df(5) nodes(50) bhazard(rate) ///
> tvc(dep5 agegrp2 agegrp3 agegrp4 agegrp5) dftvc(3)
```

Appendix D

Example code to fit the combined analytic/numeric approach with cluster robust standard errors

```
. stset rec, enter(start) f(event=1) id(id) exit(time .)
. //generate binary event (strata) indicators
. tab strata, gen(st)
. strcs group size number st2 st3 st4, df(4) tvc(st2 st3 st4) dftvc(2) vce(cluster id)
```

Acknowledgement

References

1. Miladinovic B, Kumar A, Mhaskar R, Kim S, Schonwetter R, Djulbegovic B. A flexible alternative to the Cox proportional hazards model for assessing the prognostic accuracy of hospice patient survival. *PLoS One* 2012; **7**(10):e47 804, doi: 10.1371/journal.pone.0047804. URL `http://dx.doi.org/10.1371/journal.pone.0047804`.
2. Reibnegger G. Modeling time in medical education research: The potential of new flexible parametric methods of survival analysis. *Creative Education* 2012; **3**(26):916–922.
3. Rooney J, Byrne S, Heverin M, Corr B, Elamin M, Staines A, Goldacre B, Hardiman O. Survival analysis of irish amyotrophic lateral sclerosis patients diagnosed from 1995-2010. *PLoS One* 2013; **8**(9):e74 733, doi:10.1371/journal.pone.0074733. URL `http://dx.doi.org/10.1371/journal.pone.0074733`.
4. Turnbull AE, Ruhl AP, Lau BM, Mendez-Tellez PA, Shanholtz CB, Needham DM. Timing of limitations in life support in acute lung injury patients: a multisite study*. *Crit Care Med* Feb 2014; **42**(2):296–302, doi:10.1097/CCM.0b013e3182a272db. URL `http://dx.doi.org/10.1097/CCM.0b013e3182a272db`.
5. Bwakura-Dangarembizi M, Kendall L, Bakeera-Kitaka S, Nahirya-Ntege P, Keishanyu R, Nathoo K, Spyer MJ, Kekitiinwa A, Lutaakome J, Mhute T, *et al.*. A randomized trial of prolonged co-trimoxazole in HIV-infected children in Africa. *N Engl J Med* Jan 2014; **370**(1):41–53.
6. Lambert PC, Dickman PW, Nelson CP, Royston P. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Stat Med* 2010; **29**(7-8):885–895, doi:10.1002/sim.3762. URL `http://dx.doi.org/10.1002/sim.3762`.
7. King NB, Harper S, Young ME. Use of relative and absolute effect measures in reporting health inequalities: structured review. *BMJ* 2012; **345**:e5774.
8. Eloranta S, Lambert PC, Sjöberg J, Andersson TML, Björkholm M, Dickman PW. Temporal trends in mortality from diseases of the circulatory system after treatment for hodgkin lymphoma: a population-based cohort study in sweden (1973 to 2006). *J Clin Oncol* Apr 2013; **31**(11):1435–1441, doi:10.1200/JCO.2012.45.2714. URL `http://dx.doi.org/10.1200/JCO.2012.45.2714`.
9. Lambert PC, Holmberg L, Sandin F, Bray F, Linklater KM, Purushotham A, Robinson D, Møller H. Quantifying differences in breast cancer survival between England and Norway. *Cancer Epidemiol* 2011; **35**(6):526–533, doi:10.1016/j.canep.2011.04.003. URL `http://dx.doi.org/10.1016/j.canep.2011.04.003`.
10. Andersson TML, Dickman PW, Eloranta S, Lambe M, Lambert PC. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Stat Med* Aug 2013; doi:10.1002/sim.5943. URL `http://dx.doi.org/10.1002/sim.5943`.
11. Latimer NR. Survival analysis for economic evaluations alongside clinical trials–extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making* 2013; **33**(6):743–754, doi:10.1177/0272989X12472398. URL `http://dx.doi.org/10.1177/0272989X12472398`.
12. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. *Stat Med* 2013; **32**(23):4118–4134, doi: 10.1002/sim.5823. URL `http://dx.doi.org/10.1002/sim.5823`.
13. Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *J Statist Comput Simulation* 2015; **85**(4):777–793, doi: 10.1080/00949655.2013.845890. URL `http://www.tandfonline.com/doi/abs/10.1080/00949655.2013.845890`.
14. Crowther MJ, Lambert PC. stgenreg: A Stata package for the general parametric analysis of survival data. *J Stat Softw* 2013; **53**(12).
15. Kooperberg C, Stone CJ, Truong YK. Hazard regression. *J Amer Statist Assoc* 1995; **90**(429):pp. 78–94. URL `http://www.jstor.org/stable/2291132`.
16. Kooperberg C, Clarkson DB. Hazard regression with interval-censored data. *Biometrics* 1997; **53**(4):1485–1494.
17. Royston P, Parmar MKB. Flexible Parametric Proportional Hazards and Proportional Odds Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects. *Stat Med* 2002; **21**(15):2175–2197.
18. Royston P, Lambert PC. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. Stata Press, 2011.
19. Gould W, Pitblado J, Poi B. *Maximum Likelihood Estimation with Stata*. 4th edition edn., Stata Press, 2010.
20. Stoer J, Burlirsch R. *Introduction to Numerical Analysis*. 3rd edn., Springer, 2002.
21. Begg CB, Schrag D. Attribution of deaths following cancer treatment. *J Natl Cancer Inst* Jul 2002; **94**(14):1044–1045.

22. Fall K, Strömberg F, Rosell J, Andrèn O, Varenhorst E, Group SERPC. Reliability of death certificates in prostate cancer patients. *Scand J Urol Nephrol* 2008; **42**(4):352–357, doi:10.1080/00365590802078583. URL `http://dx.doi.org/10.1080/00365590802078583`.

23. Bhaskaran K, Hamouda O, Sannes M, Boufassa F, Johnson AM, Lambert PC, Porter K, CASCADE Collaboration. Changes in the risk of death after HIV seroconversion compared with mortality in the general population. *JAMA* 2008; **300**(1):51–59, doi:10.1001/jama.300.1.51. URL `http://dx.doi.org/10.1001/jama.300.1.51`.

24. Nelson CP, Lambert PC, Squire IB, Jones DR. Flexible Parametric Models for Relative Survival, with Application in Coronary Heart Disease. *Stat Med* 2007; **26**(30):5486–5498.

25. Coleman MP, Babb P, Damiecki P, Grosclaude P, Honjo S, Jones J, Knerer G, Pitard A, Quinn M, Sloggett A, *et al.*. *Cancer Survival Trends in England and Wales, 1971–1995: Deprivation and NHS Region*. No. 61 in Studies in Medical and Population Subjects, London: The Stationery Office, 1999.

26. Bolard P, Quantin C, Abrahamowicz M, Esteve J, Giorgi R, Chadha-Boreham H, Binquet C, Faivre J. Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. *J Cancer Epidemiol Prev* 2002; **7**(3):113–122.

27. Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, Faivre J. A relative survival regression model using B-spline functions to model non-proportional hazards. *Stat Med* 2003; **22**(17):2767–2784, doi:10.1002/sim.1484. URL `http://dx.doi.org/10.1002/sim.1484`.

28. Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. *Stat Med* 2004; **23**(1):51–64, doi:10.1002/sim.1597. URL `http://dx.doi.org/10.1002/sim.1597`.

29. Remontet L, Bossard N, Belot A, Estève J, M FRANCI. An overall strategy based on regression models to estimate relative survival and model the effects of prognostic factors in cancer survival studies. *Stat Med* 2007; **26**(10):2214–2228.

30. Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1967; 221–233.

31. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 1980; **48**(4):817–838.

32. White H. Maximum likelihood estimation of misspecified models. *Econometrica* 1982; **50**:1–25.

33. Rogers WH. sg17: Regression standard errors in clustered samples. *Stata Tech Bull* 1993; **13**:19–23.

34. Durrleman S, Simon R. Flexible Regression Models with Cubic Splines. *Stat Med* 1989; **8**(5):551–561.

35. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**:373–379.

36. Therneau TM, Grambsch PM. *Modelling Survival Data: extending the Cox model*. Springer, 2000.

37. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *J Amer Statist Assoc* 1989; **84**(408):1074–1078.

38. Berry SM, Berry DA, Natarajan K, Lin C, Hennekens CH, Belder R. Bayesian survival analysis with nonproportional hazards: Metanalysis of combination pravastatin-aspirin. *Journal of the American Statistical Association* 2004; **99**(465):pp. 36–44.

39. Han G, Schell MJ, Kim J. Improved survival modeling in cancer research using a reduced piecewise exponential approach. *Stat Med* 2014; **33**(1):59–73, doi:10.1002/sim.5915. URL `http://dx.doi.org/10.1002/sim.5915`.