

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

# Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and IPD meta-analysis

Michael J. Crowther<sup>1\*†</sup>, Maxime P. Look<sup>2</sup> and Richard D. Riley<sup>3</sup>

Multilevel mixed effects survival models are used in the analysis of clustered survival data, such as repeated events, multi-centre clinical trials, and Individual Participant Data (IPD) meta-analyses, to investigate heterogeneity in baseline risk and covariate effects. In this paper we extend parametric frailty models including the exponential, Weibull, and Gompertz proportional hazards models, and the log logistic, log normal and generalised gamma accelerated failure time models, to allow any number of normally distributed random effects. Furthermore, we extend the flexible parametric survival model of Royston and Parmar, modelled on the log cumulative hazard scale using restricted cubic splines, to include random effects, whilst also allowing for non-proportional hazards (time-dependent effects). Maximum likelihood is used to estimate the models utilising adaptive or non-adaptive Gauss-Hermite quadrature. The methods are evaluated through simulation studies representing clinically plausible scenarios of a multi-centre trial and IPD meta-analysis, showing good performance of the estimation method. The flexible parametric mixed effects model is illustrated using a dataset of patients with kidney disease and repeated times to infection, and an IPD meta-analysis of prognostic factor studies in patients with breast cancer. User-friendly Stata software is provided to implement the methods. Copyright © 0000 John Wiley & Sons, Ltd.

**Keywords:** survival analysis, mixed effects, flexible parametric models, adaptive Gauss-Hermite quadrature

## 1. Introduction

The occurrence of clustered survival data is commonplace in medical research, where event times are clustered within groups of the same or similar individuals, which means event times from the same group are likely to be correlated. In this paper we are interested in situations where a researcher wants to make an average inference across all clusters whilst accounting for any heterogeneity between clusters. For example, in an Individual Participant Data

<sup>1</sup> University of Leicester, Department of Health Sciences, Adrian Building, University Road, Leicester LE1 7RH, UK.

<sup>2</sup> Department of Medical Oncology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, the Netherlands.

<sup>3</sup> University of Birmingham, School of Health and Population Sciences, Birmingham B15 2TT, UK.

\* Correspondence to: Michael J. Crowther. University of Leicester, Department of Health Sciences, Adrian Building, University Road, Leicester LE1 7RH, UK.

† E-mail: michael.crowther@le.ac.uk

(IPD) meta-analysis, individuals are clustered within different studies, and researchers are interested in synthesising the IPD to estimate the average (treatment) effect across studies and the between-study heterogeneity in the effect [1, 2, 3, 4]. This form of meta-analysis is growing in its use for survival modelling as it avoids reliance on published results, allows for the adjustment of confounders, and the investigation of non-proportional hazards [4, 5]. A second area in which clustered survival data occurs is in multi-centre clinical trials [6]. Although random treatment effects are rarely used in this context, due to greater control over inclusion/exclusion criteria, mixed models can be used to explore between-centre heterogeneity in the treatment effect. Some authors have considered centre and treatment-by-centre effects to be random in this context [7, 8, 9]. A further common area in which clustered survival data occurs is recurrent event data. Patients can experience an event of interest multiple times throughout the follow-up period, and the inherent correlation between events within patients can be accounted for by using a frailty term [10]. In this article we refer to a frailty model meaning that with only a random intercept, to distinguish that of a mixed effects model, which can have multiple (possibly correlated) random effects, of which a frailty model is a special case.

Frailty survival models have received great attention in the literature, with a variety of estimation methods and frailty distributions being proposed. The extensive *frailtypack* in R uses non-parametric penalized likelihood estimation for Cox based frailty models [11], and can provide smooth estimates of the baseline hazard function using cubic M-splines, or alternatively a Weibull or piecewise exponential hazard can be assumed. Models available include a standard shared frailty model [12], a nested frailty model [13], joint frailty [14], and a model with two random effects [15]. The two random effect model of Rondeau et al. (2008) allows for two normally distributed (possibly correlated) random effects (i.e. a random intercept and coefficient), which was applied to the setting of an IPD meta-analysis, investigating heterogeneity in baseline risk and the treatment effect. Further estimation techniques include use of an EM algorithm to estimate both frailty Cox models, mixed effect Cox models [16, 17, 18], and partial penalised likelihood [19], available in R in the *coxph* and *coxme* packages [20]. Finally, Tudur-Smith et al. (2005) and Bowden et al. (2011) have considered the use of restricted maximum likelihood within the random effect Cox model framework, and provided SAS and R code, respectively, to implement the methodology [2, 3].

Despite the Cox model remaining the most popular survival model choice [21], there is growing interest in parametric survival models [22]. Through a parametric approach, we can directly model the baseline hazard function to obtain absolute measures of risk. In particular, there is growing use of the Royston-Parmar flexible parametric survival model, which uses restricted cubic splines to model the log baseline cumulative hazard function, providing a highly flexible framework to capture complex underlying hazard functions [23, 24, 25, 26]. Many authors have shown the benefits of undertaking a flexible parametric analysis, as hazard ratio estimates closely match those from a Cox model, whilst gaining the advantages of undertaking a parametric approach, where in particular, the flexible modelling of the baseline hazard function and of time-dependent effects are particularly appealing aspects of the modelling framework [27, 28]. Knowledge of the baseline hazard or survival allows absolute risk predictions over time (e.g. in prognostic models), and enables hazards ratios to be translated back to the absolute scale (e.g. useful for calculating the number needed to treat).

Liu and Huang proposed the use of Gaussian quadrature (both adaptive and non-adaptive) for estimation in proportional hazards models with a frailty, and a piecewise constant baseline hazard function [29]. They illustrated, through simulation and application, the excellent performance of adaptive Gaussian quadrature in this framework, showing benefits over more complex and difficult to implement estimation methods, such as penalised partial likelihood. Kong et al. [30] illustrated this approach further, describing standard proportional hazard choices such as the Weibull and Gompertz distributions, and accelerated failure time models such as log-logistic and log-normal. Crowther et al. (2012) extended this approach by using mixed effects Poisson regression for the one-stage meta-analysis of IPD [4]. However, assuming a piecewise baseline hazard (or piecewise function for time-dependent effects), can limit the flexibility for both modelling and prediction, and is also potentially computationally intensive due to splitting of the data into time intervals.

In this article, we extend parametric frailty models to incorporate any number of random effects, allowing random coefficients to be included in a parametric survival analysis, estimated using adaptive or non-adaptive Gauss-Hermite quadrature. In Section 2 we describe parametric distributions including the exponential, Weibull and Gompertz proportional hazards models, and the log logistic, log normal and generalised gamma accelerated failure time models. We then extend the flexible parametric survival model of Royston and Parmar, allowing for mixed effects, which provides a highly flexible and fully parametric modelling framework to capture complex baseline hazard functions, and also extend to non-proportional hazards. In Section 3, we conduct two simulation studies to evaluate the estimation routine under a set of realistic scenarios; the first representing a multi-centre trial setting, and the second an IPD meta-analysis. In Section 4, we apply the flexible parametric mixed effects model to two clinical datasets; the first in patients with kidney disease measuring the time to repeated infection, and the second an IPD meta-analysis of prognostic studies in patients with breast cancer. Finally, we conclude the paper with a

discussion in Section 5.

## 2. Mixed effects parametric survival models

We begin with some notation. We define  $i = 1, \dots, N$  clusters (e.g. trials or centres), with each cluster having  $j = 1, \dots, n_i$  patients. Let  $S_{ij}$  be the true survival time of the  $j^{\text{th}}$  patient in the  $i^{\text{th}}$  cluster,  $T_{ij} = \min(S_{ij}, C_{ij})$  the observed survival time, with  $C_{ij}$  the censoring time. Define an event indicator  $d_{ij}$ , which takes the value of 1 if  $S_{ij} \leq C_{ij}$  and 0 otherwise.

### 2.1. Proportional hazards parametric survival models

We define the proportional hazards mixed effect survival model

$$\log [h_{ij}(t)] = \log [h_0(t)] + \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_i^T \mathbf{b}_i \quad (1)$$

where  $h_0(t)$  is the baseline hazard function of a parametric distribution, such as the exponential, Weibull or Gompertz. We define design matrices  $\mathbf{X}_{ij}$  and  $\mathbf{Z}_i$  for the fixed ( $\boldsymbol{\beta}$ ) and random ( $\mathbf{b}_i$ ) effects, respectively. We assume  $\mathbf{b}_i \sim \text{MVN}(\mathbf{0}, \mathbf{V})$ . The fixed effect design matrix,  $\mathbf{X}_{ij}$ , has dimensions  $1 \times n_f$ , where  $n_f$  is the number of covariates included as fixed effects. The random effect design matrix,  $\mathbf{Z}_i$ , has dimensions  $1 \times n_r$ , where  $n_r$  is the number of random effects.  $\mathbf{Z}_i$  can potentially contain covariates and a vector of 1's to allow a random intercept. When  $\mathbf{Z} = \mathbf{1}$ , Equation (1) reduces to a standard frailty proportional hazards model, with only a single random effect on the baseline hazard function,  $h_0(t)$ . Accelerated failure time mixed effect models are described in Appendix A.

### 2.2. Flexible parametric model

A more flexible, alternative method to commonly used parametric distributions is the flexible parametric model of Royston and Parmar [23]. This is modelled on the log cumulative hazard scale, and we can incorporate mixed effects as follows

$$\log [H_{ij}(t)] = \log [H_0(t)] + \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_i^T \mathbf{b}_i \quad (2)$$

where  $H_0(t)$  is the cumulative baseline hazard function. The spline basis for this specification is derived from the log cumulative hazard function of a Weibull proportional hazards model. The linear relationship with log time is relaxed through the use of restricted cubic splines. The restricted nature of the splines forces the function to be linear beyond the boundary knots,  $k_{\min}$  and  $k_{\max}$ . The default choice for the boundary knots are the maximum and minimum of the uncensored survival times. The spline functions are forced to join at the  $m$  internal knots,  $k_1, \dots, k_m$ , with continuous first and second derivatives. Internal knot locations are generally based on centiles of the uncensored survival times. Further details can be found in Royston and Parmar [23] and Lambert and Royston [31]. We can therefore write a restricted cubic spline function of  $\log(t)$ , with knot vector  $\mathbf{k}_0$ , letting  $u = \log(t)$

$$s(u|\boldsymbol{\gamma}, \mathbf{k}_0) = \gamma_0 + \gamma_1 v_1 + \gamma_2 v_2 + \dots + \gamma_{m+1} v_{m+1} \quad (3)$$

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

$$v_1 = u \quad (4)$$

$$v_j = (u - k_j)_+^3 - \lambda_j (u - k_{\min})_+^3 - (1 - \lambda_j)(u - k_{\max})_+^3 \quad (5)$$

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

$$\lambda_j = \frac{k_{\max} - k_j}{k_{\max} - k_{\min}} \quad (6)$$

This is now substituted for the log cumulative baseline hazard in Equation (2).

$$\log [H_{ij}(t)] = \eta_{ij} = s(\log(t)|\boldsymbol{\gamma}, \mathbf{k}_0) + \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_i^T \mathbf{b}_i \quad (7)$$

We can now transform to the hazard and survival scales

$$h_{ij}(t) = \left[ \frac{1}{t} \frac{ds(\log(t)|\boldsymbol{\gamma}, \mathbf{k}_0)}{d \log(t)} \right] \exp(\eta_{ij}), \quad S_{ij}(t) = \exp[-\exp(\eta_{ij})] \quad (8)$$

Given the fully parametric nature of the model specification, the derivatives of the spline function required in the definition of the hazard function are easily calculated. Knot location by default is based on the centiles of the uncensored survival times, which has recently been shown to provide excellent approximations for complex hazard functions, with parameter estimates robust to knot location [32].

Under Equation (2) we assume proportional cumulative hazards; however, this in fact implies proportional hazards, as in the models described in Section 2.1.

*2.2.1. Non-proportional hazards* The occurrence of non-proportional hazards (or time-dependent effects) is common in the analysis of survival data. For example, treatment effects can vary over time [33], and in registry based studies where follow-up is often over many years, covariate effects can be found to vary in magnitude over the duration of follow-up [25].

Within the flexible parametric framework, non-proportional cumulative hazards can be incorporated by interacting covariates with spline functions of log time and including them in the linear predictor [31]. Equation (7) becomes

$$\log [H_{ij}(t)] = s(\log(t)|\boldsymbol{\gamma}, \mathbf{k}_0) + \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_i^T \mathbf{b}_i + \sum_{p=1}^P s(\log(t)|\boldsymbol{\delta}_p, \mathbf{k}_p) X_{ijp} \quad (9)$$

where  $X_{ijp}$  is the  $p^{\text{th}}$  covariate interacted with spline function  $s(\log(t)|\boldsymbol{\delta}_p, \mathbf{k}_p)$ , with associated coefficient vector  $\boldsymbol{\delta}_p$ . The number of spline variables used for each time-dependent effect depends on the number of knots,  $\mathbf{k}_p$ . Often, fewer knots can be used than chosen for the baseline cumulative hazard function,  $\mathbf{k}_0$ .

### 2.3. Likelihood and estimation

We can now define the likelihood for the  $i^{\text{th}}$  cluster under the mixed effects survival modelling framework

$$L_i = \int_{-\infty}^{\infty} \left[ \prod_{j=1}^{n_i} p(T_{ij}, d_{ij} | \mathbf{b}_i, \boldsymbol{\theta}) \right] p(\mathbf{b}_i | \boldsymbol{\theta}) d\mathbf{b}_i \quad (10)$$

with parameter vector,  $\boldsymbol{\theta}$ . Under a hazard scale parametric model

$$p(T_{ij}, d_{ij} | \mathbf{b}_i, \boldsymbol{\theta}) = h(T_{ij})^{d_{ij}} \exp \left[ - \int_0^{T_{ij}} h(T_{ij}) \right] \quad (11)$$

with  $h(\cdot)$  defined as the exponential of Equation (1). Under the flexible parametric survival model (assuming proportional hazards)

$$p(T_{ij}, d_{ij} | \mathbf{b}_i, \boldsymbol{\theta}) = \left[ \left\{ \frac{1}{T_{ij}} \frac{ds(\log(T_{ij})|\boldsymbol{\gamma}, \mathbf{k}_0)}{d \log(T_{ij})} \right\} \exp(\eta_{ij}) \right]^{d_{ij}} \exp \{-\exp(\eta_{ij})\} \quad (12)$$

Finally, we assume the random effects follow a multivariate normal distribution

$$p(\mathbf{b}_i | \boldsymbol{\theta}) = (2\pi|V|)^{-q/2} \exp \left\{ - \frac{\mathbf{b}_i' V^{-1} \mathbf{b}_i}{2} \right\}, \quad (13)$$

with variance-covariance matrix,  $V$ , and  $q$  is the number of random effects. The (possibly multi-dimensional) integral in Equation (10) is analytically intractable, requiring numerical techniques to evaluate. In this article, we propose to use  $M$ -point non-adaptive or fully adaptive Gauss-Hermite quadrature [29, 34, 35]. Note that it is important to use an increasing value for  $M$  to check the consistency of results and to ensure suitable convergence has been achieved. This is described further in Section 3.

As noted in Liu and Huang [29], estimation of frailty survival models require reasonable starting values for efficient convergence of the estimation process. We adopt a two-stage estimation process as the default choice. To begin, starting values for the fixed effects are obtained from fitting the standard fixed effect parametric survival model, and values of 1 and 0 are taken for variances and covariances, respectively, for any random effects. The first stage of the estimation is then undertaken to refine the starting values, using non-adaptive Gauss-Hermite quadrature, and proceeds for 2 full Newton-Raphson iterations. The current parameter estimates are then taken as starting values for the second stage of the estimation process, switching to fully adaptive Gauss-Hermite quadrature, which proceeds until standard convergence criteria are met. The first and second derivatives are estimated numerically,

as implemented in the `m1` command in Stata [36]. We obtain variance estimates using the inverse of the negative Hessian matrix evaluated at the maximum likelihood estimates.

In the associated Stata software, four choices are available for the variance-covariance structure of the random effects, namely; independent, identity, unstructured and exchangeable.

*2.3.1. A note on the impact of the number of participants per cluster* It is important to note that when clusters are particularly large, i.e. many participants within a specific cluster, given that the numerical integration must be conducted on the likelihood function, rather than the log-likelihood function, with contributions then combined at the panel level, some re-scaling of the timescale may be required to ensure model fit. For example, in the analysis described in Section 4.2, where one cluster contained 36.61% of all patients (2722 of 7435), the timescale must be divided by a large number (e.g. 500) to allow model convergence.

#### 2.4. Stata software

The methodology described above is implemented as user-friendly Stata software, which can be downloaded within Stata by typing `ssc install stmixed`. All the models described above are available, with adaptive Gauss-Hermite quadrature the default numerical integration technique. A range of prediction options are also available, including empirical Bayes estimates of the random effects.

### 3. Simulation studies

In this section we conduct two simulation studies to assess the performance of adaptive and non-adaptive quadrature to estimate mixed effects survival models. As is often the case with methods that use numerical integration, estimation difficulties can arise with challenging datasets. To accommodate this within the simulation studies we have 5 levels of model fitting:

1. Model 1: 5-point adaptive quadrature
2. Model 2: 11-point adaptive quadrature
3. Model 3: 21-point adaptive quadrature
4. Model 4: 15-point non-adaptive quadrature
5. Model 5: 35-point non-adaptive quadrature

Each successive model is applied only if the previous fails to converge.

#### 3.1. Simulation study 1: Proportional hazards with a frailty

Liu and Huang [29] illustrated excellent performance of adaptive Gauss-Hermite quadrature in their simulation study, but only chose one frailty standard deviation of 1 with a normally distributed frailty. Here we replicate one of their simulation scenarios, but vary the magnitude of the frailty standard deviation. We assume a multi-centre trial scenario, with 100 centres, and 6 patients in each centre. A binary centre level covariate  $X_{1i}$  (such as treatment if individual centres were randomised to a single treatment group) is generated from  $\text{Bin}(1, 0.5)$ , and a patient specific covariate (such as a biomarker level),  $X_{2ij}$  is generated from  $\text{U}(0, 1)$ , with associated fixed effects of  $\{1, -1\}$ , respectively. We therefore simulate from the following model

$$h_{ij}(t) = \lambda \gamma t^{\gamma-1} \exp [b_{0i} + \beta_1 X_{1i} + \beta_2 X_{2ij}] \quad (14)$$

assuming a Weibull baseline hazard with scale  $\lambda = 1$  and shape  $\gamma = 2$ . We have a hazard ratio of  $\exp(\beta_1) = 2.718$  for a one-unit increase in  $X_{1i}$ , assuming  $X_{2ij}$  is held constant, and  $\exp(\beta_2) = 0.368$ , our hazard ratio for a one-unit increase in  $X_{2ij}$  given  $X_{1i}$  is held constant. Censoring times are generated from  $\text{U}(0, 2)$ . A normal frailty is incorporated from  $b_{0i} \sim N(0, \sigma^2)$ , with  $\sigma = \{0.2, 0.5, 1\}$ . In each scenario we conduct 1000 replications. A Weibull survival model with frailty is applied to each simulated dataset.

Results are presented in Table 1, including the number of simulations that converged at each model fitting stage, and in total. Bias, percentage bias and coverage are calculated on the parameterised scales, using estimates of the log hazard ratios,  $\beta_1$  and  $\beta_2$ , and the log of the scale and shape parameters of the baseline hazard, and the heterogeneity standard deviation, i.e.  $\log(\lambda)$ ,  $\log(\gamma)$  and  $\log(\sigma)$ . When  $\sigma = 1$ , we observe entirely consistent results to that of Liu and Huang [29], showing excellent performance of the estimation method, with minimal bias in all parameter estimates and coverage probabilities close to the optimum of 95%. When  $\sigma = 0.5$ , we again find minimal bias and optimum coverage probabilities across parameters, with only some minor underestimation of the frailty

standard deviation,  $\sigma$ . Both baseline scale and shape parameters,  $\lambda$  and  $\gamma$ , have been captured with minimal bias and optimum coverage probabilities, across scenarios 1 and 2. However, when  $\sigma = 0.2$ , 92 of the 1000 simulations failed to converge across all 5 of the models applied. Of the 908 that did converge, we observe minimal bias in the log hazard ratios  $\beta_1$  and  $\beta_2$ , and coverage levels of 86.5%, indicating some underestimation of standard errors. In particular, we observe substantial bias and percentage bias in estimates of  $\log(\sigma)$ ; however, this is because the mean calculation of bias is skewed due to some large negative estimates (or equivalently estimates of  $\sigma$  close to 0). Using the median to summarise estimates of  $\sigma$  is much more informative in this case, due to  $\sigma$  being bounded by zero, in which case we observe a median estimate of 0.191, close to the true value of 0.2. This result should be interpreted with caution given the amount of non-convergence.

**Table 1.** Simulation study 1: Weibull baseline with normal frailty.

Parameter	Estimate	Bias	% bias	CP	C1	C2	C3	C4	C5	Total Conv.
<i>Scenario 1</i>					1000	0	0	0	0	1000
$\beta_1 = 1$	1.006*	0.006	0.6	94.4						
$\beta_2 = -1$	-1.016*	-0.016	1.6	94.4						
$\lambda = 1$	0.999†	-0.012	.	94.9						
$\gamma = 2$	2.007†	0.004	0.6	94.7						
$\sigma = 1$	0.987†	-0.018	.	94.0						
<i>Scenario 2</i>					990	10	0	0	0	1000
$\beta_1 = 1$	1.003*	0.003	0.3	94.8						
$\beta_2 = -1$	-1.016*	-0.016	1.6	94.4						
$\lambda = 1$	1.006†	0.005	.	95.5						
$\gamma = 2$	2.007†	0.003	0.4	95.3						
$\sigma = 0.5$	0.486†	-0.059	8.5	96.7						
<i>Scenario 3</i>					474	227	9	196	2	908
$\beta_1 = 1$	1.010*	0.010	1.0	86.5						
$\beta_2 = -1$	-1.019*	-0.019	1.9	86.5						
$\lambda = 1$	1.010†	0.007	.	85.8						
$\gamma = 2$	2.012†	0.007	1.0	86.8						
$\sigma = 0.2$	0.191†	-1.700	105.7	83.3						

\* - mean, † - median, CP - coverage probability

C# indicates the number of simulations that converged

Bias, % bias and CP are calculated on the parameterised scale ( $\log(\lambda)$ ,  $\log(\gamma)$ ,  $\log(\sigma)$ ), and on simulations that converged.

### 3.2. Simulation study 2: Proportional trial effects with a random treatment effect

We now conduct a simulation study to assess the performance of the estimation method in scenarios representing IPD meta-analyses of survival data, incorporating a random coefficient, i.e. a treatment effect with heterogeneity. It is well-known that maximum likelihood estimation leads to downwardly biased estimates of variance parameters in small samples which, in a meta-analysis, is when the number of studies is small. This has been shown in other simulation studies, such as mixed-effects logistic regression and Poisson regression models [4, 37]. We investigate this further by varying the magnitude of the heterogeneity variance, and applying mixed effect Weibull proportional hazards models. We simulate from the following model

$$h_{ij}(t) = \lambda \gamma t^{\gamma-1} \exp[\beta_{0i} + (\beta_1 + b_{1i})X_{1ij}] \tag{15}$$

assuming a Weibull distribution, with shape and scale parameters of  $\gamma = 1.276$  and  $\lambda = 3.121$ , respectively. We incorporate proportional trial effects,  $\beta_{0i}$ , for  $i = 2, \dots, n$ , drawn from  $N(0, 0.25^2)$ , with  $\beta_{01}$  constrained to be zero for the reference trial, and therefore  $\exp(\beta_{0i})$  represents the proportional effect on the baseline hazard due to the  $i^{th}$  trial, where  $i = 2, \dots, n$ . We generate a binary covariate,  $X_{1ij} \sim \text{Bin}(1, 0.5)$ , to represent treatment group, and

where  $\beta_1$  represents the mean log hazard ratio for a population of treatment effects, with  $b_{1i}$  the deviation of the log hazard ratio of the  $i^{th}$  trial from the population mean. We therefore assume  $(\beta_1 + b_{1i}) \sim N(-0.663, \tau^2)$ , with  $\tau = \{0.25, 0.5, 1\}$ . Administrative censoring is applied at 0.24 units. We assume 15 trials with 500 patients in each. The number of patients represents an approximate average trial size for the example described below in Section 4.2, which also informs the random treatment effect mean and standard deviation, and baseline parameters for the Weibull distribution, to provide plausible simulation scenarios. Equation (15) is then applied to each simulated dataset.

**Table 2.** Simulation study 2: Weibull baseline with random treatment effect and proportional trial effects.

Parameter	Estimate	Bias	% bias	CP	C1	C2	C3	C4	C5	Total Conv.
<i>Scenario 1</i>					961	19	11	9	0	1000
$\beta_1 = -0.663$	-0.649*	0.014	-2.1	91.2						
$\lambda = 3.121$	3.121 <sup>†</sup>	-0.005	-0.4	94.2						
$\gamma = 1.276$	1.278 <sup>†</sup>	0.001	0.4	95.7						
$\tau = 1$	0.937 <sup>†</sup>	-0.081	.	89.9						
<i>Scenario 2</i>					872	115	13	0	0	1000
$\beta_1 = -0.663$	-0.654*	0.009	-1.4	91.0						
$\lambda = 3.121$	3.125 <sup>†</sup>	-0.004	-0.4	93.9						
$\gamma = 1.276$	1.277 <sup>†</sup>	0.001	0.4	96.0						
$\tau = 0.5$	0.460 <sup>†</sup>	-0.116	16.7	91.0						
<i>Scenario 3</i>					194	466	246	33	6	945
$\beta_1 = -0.663$	-0.655*	0.008	-1.2	84.3						
$\lambda = 3.121$	3.115 <sup>†</sup>	0.008	0.3	88.3						
$\gamma = 1.276$	1.277 <sup>†</sup>	0.001	0.1	90.7						
$\tau = 0.25$	0.208 <sup>†</sup>	-0.898	64.8	93.7						

\* - mean, † - median, CP - coverage probability

C# indicates the number of simulations that converged

Bias, % bias and CP are calculated on the parameterised scale ( $\log(\lambda)$ ,  $\log(\gamma)$ ,  $\log(\sigma)$ ), and on simulations that converged.

Results are presented in Table 2, including the number of simulations which converged at each of the 5 nested model stages. Once again bias, percentage bias and coverage are calculated on the parameterised scales, using estimates of the log hazard ratio,  $\beta_1$ , and the log of the scale and shape parameters of the baseline hazard, and the heterogeneity standard deviation, i.e.  $\log(\lambda)$ ,  $\log(\gamma)$  and  $\log(\tau)$ . When  $\tau = 1$  or  $\tau = 0.5$ , we observe very good performance of the estimation method, with only minor underestimation of the heterogeneity standard deviation,  $\tau$ . Estimates of the mean treatment effect,  $\beta_1$  are generally unbiased, with slightly sub-optimum coverage probabilities of 91.2% and 91.0% in scenarios 1 and 2, respectively. Estimates of baseline parameters,  $\lambda$  and  $\gamma$ , were essentially unbiased and had optimum coverage probabilities. However, when  $\tau = 0.25$ , some issues with convergence were observed, with 55 of the 1000 simulations failing to converge across all five of the model fitting stages. Of those that did converge, we generally observe unbiased estimates of the average treatment effect, but moderate underestimation of the heterogeneity standard deviation (median of estimates = 0.208) was seen in the simulations that did converge. As in simulation study 1, mean bias and percentage bias of  $\log(\tau)$  was substantial due to some large negative estimates, on the log scale, equivalent to some estimates of  $\tau$  close to 0. Of course this result must be interpreted with caution given the moderate amount of non-convergence.

## 4. Example applications

### 4.1. Kidney data - repeated times to infection

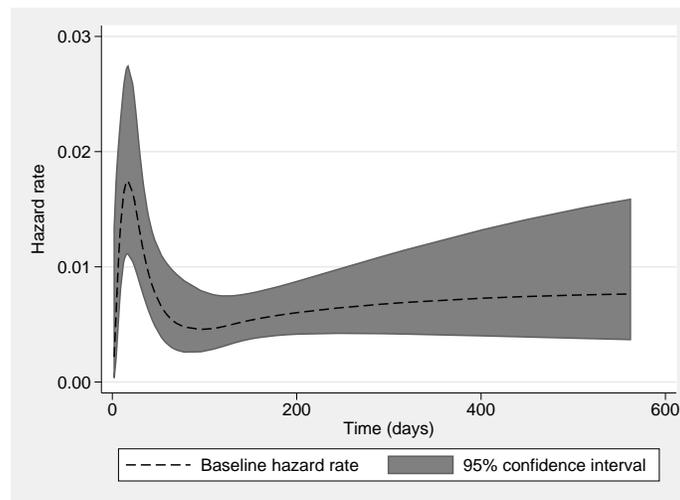
In this section we consider a re-analysis of a dataset given in McGilchrist and Aisbett (1991) [19]. The dataset consists of 38 patients with kidney disease, where the event of interest is infection at the catheter insertion point.

Each patient has two possible recurrence times, recorded from initial insertion. A total of 58 failures were observed. The clinical question of interest is to examine the effect of age and sex on the hazard of infection, and accounting for potentially recurrent events for each individual (in this case the clusters are the individual patients themselves). Here we wish to fit a flexible parametric frailty model. In preliminary modelling, with no covariates or frailty, we can use the AIC to guide the choice of number of degrees of freedom required to capture the baseline hazard function [32], with results presented in Table 3.

**Table 3.** Model fit criteria across varying degrees of freedom for the baseline hazard function using a flexible parametric frailty model.

Baseline degrees of freedom	log-likelihood	AIC
1	-107.469	218.938
2	-105.672	217.345
3	-101.872	211.745
4	-101.846	213.691
5	-101.445	214.889
6	-100.017	214.034
7	-99.727	215.454
8	-99.632	217.264
9	-98.306	216.612

From Table 3, we see that the AIC selects 3 degrees of freedom as the optimum choice for the baseline hazard function, with the estimated function shown in Figure 1. Figure 1 indicates the presence of 2 turning points in the underlying hazard function, which would clearly be missed if fitting a standard parametric frailty model such as the Weibull. We can now fit a frailty flexible parametric survival model



**Figure 1.** Estimated baseline hazard function with 3 degrees of freedom.

$$h_{ij}(t) = h_0(t) \exp(b_{0i}) \tag{16}$$

where  $b_{0i} \sim N(0, \sigma^2)$ . To give an indication of the heterogeneity in the underlying hazard across individuals, we can calculate patient-specific predictions of the hazard function, including empirical Bayes predictions of the frailty effect [35], shown in Figure 2. Figure 2 shows substantial variation in the baseline hazard across patients. We can now incorporate covariates under the following model

$$h_{ij}(t) = h_0(t) \exp(b_{0i} + \beta_1 X_{1ij} + \beta_2 X_{2ij}) \tag{17}$$

adjusting for age (years),  $X_{1ij}$ , and sex (male as the reference group),  $X_{2ij}$ , with associated log hazard ratios,  $\beta_1$  and  $\beta_2$ , respectively. Results presented in Table 4, where results are also presented from a Cox frailty model for comparison.

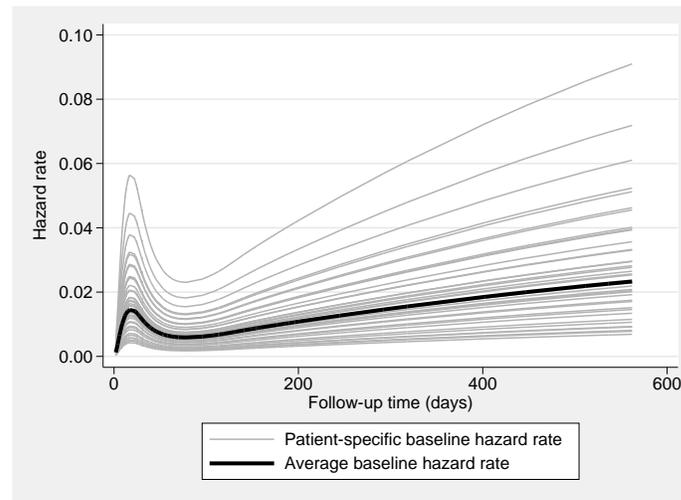


Figure 2. Patient-specific estimates of the baseline hazard function.

Table 4. Results from the flexible parametric frailty model with 3 degrees of freedom for the baseline hazard function, and a Cox frailty model.

Covariate	Flexible parametric frailty model			Cox frailty model		
	Hazard ratio	95% CI		Hazard ratio	95% CI	
Age (years)	1.007	0.982	1.034	1.005	0.981	1.030
Female	0.231	0.088	0.605	0.252	0.107	0.600
$\sigma$	0.800	0.415	1.542	0.771	-	-

From Table 4, results from the flexible parametric frailty model show a statistically significant difference between females and males, with a hazard ratio of 0.231 (95% CI: 0.088, 0.605) for females compared to males of the same age. We also find a non-statistically significant hazard ratio of 1.007 (95% CI: 0.982, 1.034) for a one-unit increase in age, assuming the same sex. We observe a heterogeneity standard deviation of 0.800 (95% CI: 0.415, 1.542), indicating that across individuals the hazard function is highly heterogeneous. It is important to establish the consistency of the numerical integration by using an increasing number of quadrature points and comparing estimates [35]. In this case, we obtained identical estimates up to 4 decimal places with 9 and 10 adaptive quadrature points.

We also compared the flexible parametric frailty model to the Cox frailty model. Results from Table 4 show similar estimates in both the hazard ratios and associated 95% confidence intervals, and the estimated frailty standard deviation. The penalised partial likelihood approach used to fit the Cox frailty model does not provide a confidence interval for frailty parameters.

One could potentially use the fitted flexible parametric model to make predictions over time for a new individual with a given age and sex, as we have the average baseline risk and the hazard ratios for age and sex. The Cox model does not give the average baseline risk, and so could not be used for this purpose.

#### 4.2. Breast cancer - IPD meta-analysis of prognostic biomarkers

We now apply the mixed effect flexible parametric survival model to an IPD meta-analysis of prognostic studies in patients with breast cancer [38]. IPD was obtained from 15 studies, with a total of 7435 patients, 2042 (27.48%) of which died.

For illustration purposes, we are interested in the effect of hormone receptor status on the hazard of death over time, coded  $-\frac{1}{2}$  for negative or unknown, and  $\frac{1}{2}$  for at least one positive. Hormone receptor status is determined by two measurements, namely estrogen receptor (ER) and progesterone receptor (PgR) dichotomised into low and high. When a random effect is placed on hormone receptor status, this group coding assumes equal variability in the log hazard rate for both groups [2].

Commonly used one-stage meta-analysis models will assume either proportional study effects, or stratify by study membership. Within a parametric modelling framework, stratifying by study membership allows the estimation of

separate baseline hazard functions for each study, which is often more plausible than assuming proportional study effects. We can assess this through preliminary fixed effect models and using model selection criteria. Applying a flexible parametric survival model to the IPD with proportional study effects, and using the AIC to select the number of degrees of freedom for the reference study (3 degrees of freedom in this case), we obtain an AIC of 11348.37. Comparing this to a flexible parametric model stratified by study, with the degrees of freedom for each study’s baseline hazard selected using the AIC, we obtain an AIC of 11342.978, indicating an improved fit when stratifying by study. The estimated baseline hazard functions are shown in Figure 3. Of course, within the mixed

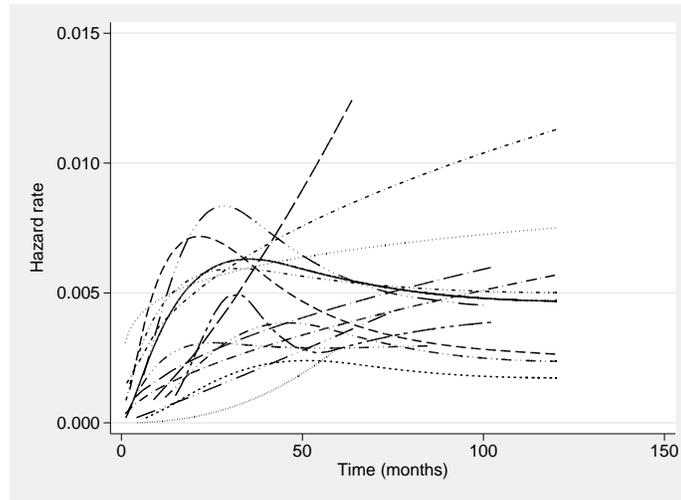


Figure 3. Estimated baseline hazard functions for the 15 trials.

effects framework, we can assume a random baseline, i.e. random study effects; however, this assumes that the studies used in the analysis are a random sample of a distribution of studies. This approach has been advocated by some authors [15, 39], particularly as it is clinically useful to obtain an estimate of the average risk (across reference groups), in which case, a parametric approach is preferable.

In our example, we proceed by stratifying by study, and investigate heterogeneity in the effect of hormone receptor status, under the following flexible parametric model.

$$h_{ij}(t) = h_{0i}(t) \exp[(\beta_1 + b_{1i})X_{1ij}], \quad \text{where } b_{1i} \sim N(0, \tau^2) \tag{18}$$

where  $h_{0i}(t)$  is the baseline hazard function for the  $i^{th}$  study, with the baseline hazard for each trial modelled using restricted cubic splines within the flexible parametric framework,  $X_{1ij}$  is hormone receptor status,  $\beta_1$  is the average log hazard ratio for a distribution of covariate effects, with  $b_{1i}$  the deviation of the  $i^{th}$  study from this average effect. For comparison, we also fit the equivalent stratified Cox mixed effect model. Results are presented in Table 5.

Table 5. Results from one-stage IPD meta-analysis, stratified by study with random hormone receptor status.

Parameter	Royston-Parmar model			Cox model		
	Estimate	95% CI		Estimate	95% CI	
$\beta_1$	-0.663	-0.854	-0.471	-0.669	-0.865	-0.473
$\exp(\beta_1)$	0.515	0.426	0.624	0.512	0.421	0.623
$\tau$	0.257	0.135	0.490	0.272	-	-

From Table 5, we observe a statistically significant average effect of hormone receptor status, with a hazard ratio of 0.515 (95% CI: 0.426, 0.624), with an estimated heterogeneity standard deviation of  $\tau = 0.257$  (95% CI: 0.135, 0.490). Consistent estimates, to 4 decimal places, were obtained with 5 and 6 adaptive quadrature points. Comparing estimates from the flexible parametric mixed effect model and the Cox mixed effect model, we observe very similar estimates of log hazard ratios and 95% confidence intervals, with similar estimates of the heterogeneity standard deviation in the effect of hormone receptor status.

It is also often of interest to adjust for patient-level covariates, particularly to see if any heterogeneity in, for example a treatment effect, is explained by covariates. Within the proposed framework, any number of covariates

can be included as fixed, or indeed random effects. To illustrate this, we now include age as a fixed effect, with results presented in Table 6. We can see that after adjusting for age, the heterogeneity in the effect of hormone

**Table 6.** Results from one-stage IPD meta-analysis, adjusting for age and hormone receptor status.

Covariate	Royston-Parmar model			Cox model		
	Hazard ratio	95% CI		Hazard ratio	95% CI	
Hormone receptor	0.493	0.409	0.596	0.490	0.404	0.595
Age (years)	1.013	1.010	1.017	1.013	1.010	1.017
$\tau$	0.249	0.130	0.476	0.264	-	-

receptor status has reduced slightly from  $\tau = 0.257$  to  $\tau = 0.249$ , and we obtain a hazard ratio of 1.013 (95% CI: 1.010, 1.017) for a one year increase in age. We once again obtained consistent estimates to 4 decimal places with 5 and 6 adaptive quadrature points. We again find very similar estimates from the stratified Cox mixed effect model.

Following a random effects meta-analysis, we can calculate a prediction interval to provide a range for the predicted parameter in a new study. For the hazard ratio of hormone receptor status, this gives a 95% prediction interval of (0.277, 0.878), calculated on the natural log scale and then exponentiated [40, 41]. As the prediction interval does not contain 1, this reveals that in at least 95% of settings hormone receptor status will have prognostic value (in the same direction). This finding is masked when ignoring heterogeneity and when focusing on only the 95% CI for the summary effect. Therefore, despite the heterogeneity, there is strong consistent evidence here that hormone receptor status is a prognostic factor for breast cancer patients across the settings covered by these studies.

Further extensions such as treatment-covariate interactions, non-proportional hazards, and multiple random effects all follow naturally within the framework and are available within the associated Stata package.

## 5. Discussion

In this article we developed methodology for the analysis of clustered survival data, incorporating mixed effects into the parametric survival analysis framework using maximum likelihood. This extends parametric frailty survival models, allowing any number of normally distributed random effects to be used, including the exponential, Weibull, and Gompertz proportional hazards models, and the log logistic, log normal and generalised gamma accelerated failure time models. Furthermore, we extended the flexible parametric survival modelling framework of Royston and Parmar (2002), modelled on the log cumulative hazard scale using restricted cubic splines, to include random effects, and also allowing for non-proportional hazards (time-dependent effects). This framework has the potential to be used in a variety of settings, such as the analysis of repeated event data, the one-stage IPD meta-analysis of survival data, the analysis of multi-centre trial data, and the development of prognostic models using data from multiple clusters.

There is growing use in parametric survival models, in particular, the class of Royston-Parmar flexible parametric models [27, 28]. By incorporating mixed effects into this framework, we demonstrate a useful modelling approach to capture complex hazard functions, often observed in real datasets. Furthermore, we extend to time-dependent effects in flexible parametric mixed effects survival models. In the example applications, we used model selection criteria, such as the AIC, to select the degrees of freedom for the baseline hazard function. This, combined with sensitivity analysis for knot location, have been shown to be highly effective in model selection [24, 26, 32]. Through the flexible parametric approach we can model the baseline and time-dependent effects in continuous time, allowing smooth predictions, both in and out of sample, compared to other piecewise approaches [4, 29].

We conducted two simulation studies in this article, to assess the performance of the estimation routine in a set of clinically realistic scenarios, including a multi-centre clinical trial, and an IPD meta-analysis. In simulation study 1, chosen to reflect a multi-centre clinical trial context, we found excellent performance in estimating the frailty variance when moderate and large values of the variance were chosen. However, when small values were chosen, some convergence issues arose, with some bias and sub-optimum coverage probabilities. Log hazard ratios and baseline parameters were essentially unbiased with coverage probabilities at approximately 95%. In simulation study 2, we simulated data to represent an IPD meta-analysis, similar to that of the prognostic study meta-analysis analysed in Section 4.2. In agreement with the findings of Crowther et al. (2012) [4], some underestimation of random effect standard deviations were observed in this simulation setting, although the magnitude of the bias appears to decrease

as the magnitude of the standard deviation increases. Baseline hazard parameters of the Weibull distributions used were also captured with minimal bias and optimum coverage probabilities.

In the application to the kidney data, we highlighted the ability of the flexible parametric frailty model to capture complex hazard functions, which detected 2 turning points in the underlying hazard function. If interest lies in measures absolute risk such as the hazard rate, which are clinically meaningful from an epidemiological perspective, it is crucial to correctly model the baseline. This is especially important for prognostic models, which need to predict absolute risk over time.

It is generally accepted that one-stage models are preferable for the meta-analysis of survival data, but are often described as computationally formidable [42]. In the example analyses shown in this article, the maximum computation time was less than 5 minutes on a standard laptop computer with 4gb of RAM, which was applied to the meta-analysis of 7435 patients, indicating this framework is entirely feasible for application to clinical datasets. Some authors have proposed to incorporate a random effect on the baseline risk, and thus not include proportional or separate baseline hazards for each study [2, 15]. Indeed, this may be preferable in situations with large numbers of studies, to speed up estimation. More importantly, it is useful if a meta-analysis can provide us with the average baseline risk across all trials (across all placebo/reference groups), which falls directly into the parametric framework described in this article. Subsequently, relative effects can be easily transferred to an absolute scale. Schmid et al. (2004) advocate this approach [39]: “The assumption of random trial effects in [2] implies that results of trials with similar measured characteristics will vary because of unmeasured characteristics. In essence, each trial is a random sample from a population of potential trials. It is more common in multicenter trials for which inference is desired about the particular centers to treat the study terms as fixed effects [29]. We would argue that the random effects formulation is more appropriate to meta-analysis because consumers wish to generalise results to future trials or to different settings. However, this view is by no means universal [30].”

A final benefit of undertaking a parametric approach is in the area of economic decision modelling [43], where survival is often required to be extrapolated to a lifetime horizon, beyond the range of the data. Through a parametric approach, uncertainty in baseline risk can be directly incorporated into the economic model.

Further work is needed to evaluate the estimation methods in clinically realistic scenarios, to investigate how well it performs for varying effect sizes, heterogeneity standard deviation sizes, number of clusters, and number of participants per cluster. Previous work has also investigated the use of restricted maximum likelihood within the Cox mixed effects framework [44, 3], which could be extended to the parametric mixed effect framework described here, with the aim to reduce the underestimation in heterogeneity parameters under a maximum likelihood approach.

To facilitate the use of the methods in practice, user friendly Stata software has been developed by the first author. All the models described in this article are available, with a variety of prediction options. Further possible extensions to the framework include allowing for higher levels of clustering, and further simulation studies to establish robustness to underlying assumptions, such as normally distributed random effects.

## Acknowledgement

Michael Crowther is funded by a National Institute for Health Research (NIHR) Doctoral Research Fellowship (DRF-2012-05-409). Maxime Look is funded by the ERC (Grant: ERC-2012-AdG-322737). Richard Riley is supported by funding from the MRC PROGnosis REsearch Strategy (PROGRESS) Partnership ([www.progress-partnership.org](http://www.progress-partnership.org)) (Medical Research Council Partnership Grant: G0902393).

## A. Appendix

### A.1. Accelerated failure time models

Mixed effects can also be incorporated into an accelerated failure time (AFT) framework. This is included in the associated Stata package, and therefore we briefly discuss them here.

*A.1.1. Log logistic* We define the survival and density functions for the log-logistic accelerated failure time model

$$S_{ij}(t) = \{1 + (\lambda_{ij}t)^{1/\gamma}\}^{-\gamma}, \quad f(t) = \frac{\lambda_{ij}^{1/\gamma} t^{1/\gamma-1}}{\gamma \{1 + (\lambda_{ij}t)^{1/\gamma}\}^2} \quad (19)$$

where

$$\log(\lambda_{ij}) = -(\mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_i^T \mathbf{b}_i) \quad (20)$$

and  $\gamma$  treated as an ancillary parameter.

*A.1.2. Log normal* We define the survival and density functions for the log-normal accelerated failure time model

$$S_{ij}(t) = 1 - \Phi\left(\frac{\log(t) - \mu_{ij}}{\sigma}\right), \quad f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\{\log(t) - \mu_{ij}\}^2\right] \quad (21)$$

where

$$\mu_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_i^T \mathbf{b}_i$$

$\Phi(z)$  is the standard normal cumulative distribution function, and  $\sigma$  treated as an ancillary parameter.

*A.1.3. Generalised gamma* Finally, we define the survival function for the generalised gamma accelerated failure time model

$$S_{ij}(t) = \begin{cases} 1 - I(\gamma, u), & \text{if } \kappa > 0 \\ 1 - \Phi(z), & \text{if } \kappa = 0 \\ I(\gamma, u), & \text{if } \kappa < 0 \end{cases} \quad (22)$$

and the density function

$$f_{ij}(t) = \begin{cases} 1 - I(\gamma, u), & \text{if } \kappa > 0 \\ 1 - \Phi(z), & \text{if } \kappa = 0 \end{cases} \quad (23)$$

where  $\gamma = |\kappa|^{-2}$ ,  $z = \text{sign}(\kappa)\{\log(t) - \mu_{ij}\}/\sigma$ ,  $u = \gamma \exp(|\kappa|z)$ ,  $\Phi(z)$  is the standard normal cumulative distribution function, and  $I(a, x)$  is the incomplete gamma function. The model is parameterised as follows

$$\mu_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \mathbf{Z}_i^T \mathbf{b}_i$$

with  $\kappa$  and  $\sigma$  treated as ancillary parameters. The generalised gamma is fairly flexible, with special cases the Weibull distribution when  $\kappa = 1$ , the exponential when  $\kappa = 1$  and  $\sigma = 1$ , and the log normal distribution when  $\kappa = 0$ .

Interpretation in the AFT framework is not as intuitive as within the proportional hazards metric. For example, if a coefficient is positive, then a one-unit increase in the covariate decelerates time, in other words it increases the expected time to failure.

*A.1.4. Likelihood* Under an accelerated failure time model we have

$$p(T_{ij}, d_{ij} | \mathbf{b}_i, \boldsymbol{\theta}) = f(T_{ij})^{d_{ij}} \times S(T_{ij})^{1-d_{ij}} \quad (24)$$

with  $f(T_{ij})$  and  $S(T_{ij})$  defined above.

## References

1. Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials* 2005; **2**(3):209–217.
2. Tudur-Smith C, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med* 2005; **24**(9):1307–1319.
3. Bowden J, Tierney JF, Simmonds M, Copas AJ, Higgins JP. Individual patient data meta-analysis of time-to-event outcomes: one-stage versus two-stage approaches for estimating the hazard ratio under a random effects model. *Research Synthesis Methods* 2011; **2**(3):150–162, doi:10.1002/jrsm.45. URL <http://dx.doi.org/10.1002/jrsm.45>.
4. Crowther MJ, Riley RD, Staessen JA, Wang J, Gueyffier F, Lambert PC. Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Med Res Methodol* 2012; **12**(34), doi:10.1186/1471-2288-12-34. URL <http://dx.doi.org/10.1186/1471-2288-12-34>.
5. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; **340**:c221.
6. Komárek A, Lesaffre E, Legrand C. Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution. *Stat Med* Dec 2007; **26**(30):5457–5472, doi:10.1002/sim.3083. URL <http://dx.doi.org/10.1002/sim.3083>.
7. Yamaguchi T, Ohashi Y, Matsuyama Y. Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials. *Stat Methods Med Res* 2002; **11**(3):221–236.
8. Legrand C, Ducrocq V, Janssen P, Sylvester R, Duchateau L. A Bayesian approach to jointly estimate centre and treatment by centre heterogeneity in a proportional hazards model. *Stat Med* 2005; **24**(24):3789–3804.
9. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. statistical principles for clinical trials. *Statistics in Medicine* 1999; **18**:1905–1942.

10. Gutierrez RG. Parametric frailty and shared frailty survival models. *The Stata Journal* 2002; **2**(1):22–44.
11. Rondeau V, Mazroui Y, Gonzalez JR. frailtypack: An R Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation. *Journal of Statistical Software* 2012; **47**(4).
12. Rondeau V, Commenges D, Joly P. Maximum penalized likelihood estimation in a gamma-frailty model. *Lifetime Data Anal Jun* 2003; **9**(2):139–153.
13. Rondeau V, Filleul L, Joly P. Nested frailty models using maximum penalized likelihood estimation. *Stat Med* Dec 2006; **25**(23):4036–4052, doi:10.1002/sim.2510. URL <http://dx.doi.org/10.1002/sim.2510>.
14. Rondeau V, Mathoulin-Pelissier S, Jacqmin-Gadda H, Brouste V, Soubeyran P. Joint frailty models for recurring events and death using maximum penalized likelihood estimation: application on cancer events. *Biostatistics* Oct 2007; **8**(4):708–721, doi: 10.1093/biostatistics/kxl043. URL <http://dx.doi.org/10.1093/biostatistics/kxl043>.
15. Rondeau V, Michiels S, Liquet B, Pignon JP. Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Stat Med* 2008; **27**(11):1894–1910.
16. Nielsen GG, Gill RD, Andersen PK, Sørensen TIA. A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* 1992; **19**(1):pp. 25–43. URL <http://www.jstor.org/stable/4616223>.
17. Guo G, Rodriguez G. Estimating a multivariate proportional hazards model for clustered data using the em algorithm, with an application to child survival in guatemala. *Journal of the American Statistical Association* 1992; **87**(420):pp. 969–976. URL <http://www.jstor.org/stable/2290633>.
18. Klein JP. Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics* 1992; **48**(3):pp. 795–806. URL <http://www.jstor.org/stable/2532345>.
19. McGilchrist CA, Aisbett CW. Regression with frailty in survival analysis. *Biometrics* Jun 1991; **47**(2):461–466.
20. Therneau TM, Grambsch PM. *Modelling Survival Data: extending the Cox model*. Springer, 2000.
21. Cox DR. Regression models and life-tables. *J Roy Statist Soc Ser B* 1972; **34**(2):187–220.
22. King NB, Harper S, Young ME. Use of relative and absolute effect measures in reporting health inequalities: structured review. *BMJ* 2012; **345**:e5774.
23. 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.
24. Royston P, Lambert PC. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. Stata Press, 2011.
25. 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>.
26. Crowther MJ, Abrams KR, Lambert PC. Flexible parametric joint modelling of longitudinal and survival data. *Stat Med* 2012; **31**(30):4456–4471, doi:10.1002/sim.5644. URL <http://dx.doi.org/10.1002/sim.5644>.
27. 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):e47804, doi: 10.1371/journal.pone.0047804. URL <http://dx.doi.org/10.1371/journal.pone.0047804>.
28. 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):e74733, doi:10.1371/journal.pone.0074733. URL <http://dx.doi.org/10.1371/journal.pone.0074733>.
29. Liu L, Huang X. The use of gaussian quadrature for estimation in frailty proportional hazards models. *Stat Med* Jun 2008; **27**(14):2665–2683, doi:10.1002/sim.3077. URL <http://dx.doi.org/10.1002/sim.3077>.
30. Kong X, Archer KJ, Moulton LH, Gray RH, Wang MC. Parametric frailty models for clustered data with arbitrary censoring: application to effect of male circumcision on hpv clearance. *BMC Med Res Methodol* 2010; **10**:40, doi:10.1186/1471-2288-10-40. URL <http://dx.doi.org/10.1186/1471-2288-10-40>.
31. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata J* 2009; **9**:265–290.
32. 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>.
33. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpaweravong P, Han B, Margono B, Ichinose Y, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009; **361**(10):947–957, doi:10.1056/NEJMoa0810699. URL <http://dx.doi.org/10.1056/NEJMoa0810699>.
34. Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J Comput Graph Statist* 1995; **4**(1):pp. 12–35.
35. Rabe-Hesketh S, Skrondal A, Pickles A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata J* 2002; **2**:1–21.
36. Gould W, Pitblado J, Poi B. *Maximum Likelihood Estimation with Stata*. 4th edition edn., Stata Press, 2010.
37. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Thompson JR. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 2007; **7**:3, doi:10.1186/1471-2288-7-3. URL <http://dx.doi.org/10.1186/1471-2288-7-3>.
38. Look MP, van Putten WLJ, Duffy MJ, Harbeck N, Christensen IJ, Thomssen C, Kates R, Spyrtos F, Fern andez M, Eppenberger-Castori S, et al. Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor pai-1 in 8377 breast cancer patients. *J Natl Cancer Inst* Jan 2002; **94**(2):116–128.
39. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol* 2004; **57**(7):683–697.
40. Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2009; **172**(1):137–159, doi:10.1111/j.1467-985X.2008.00552.x.
41. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ* 2011; **342**:d549, doi:10.1136/bmj.d549.
42. Thompson S, Kaptoge S, White I, Wood A, Perry P, Danesh J, Collaboration TERF. Statistical methods for the time-to-event

- analysis of individual participant data from multiple epidemiological studies. *International Journal of Epidemiology* 2010; **39**:1345–1359, doi:10.1093/ije/dyq063. URL <http://ije.oxfordjournals.org/content/early/2010/05/03/ije.dyq063.abstract>.
43. 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>.
44. McGilchrist CA. REML estimation for survival models with frailty. *Biometrics* Mar 1993; **49**(1):221–225.