Parametric accelerated failure time models with random effects and an application to kidney transplant survival

Philippe Lambert[1,2]*, Dave Collett[3], Alan Kimber[4] and Rachel Johnson[3]

[1] *Institut de Statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium*

[2] *Unité d'épidémiologie, biostatistique et méthodes opérationnelles, Faculté de Médecine, Université catholique de Louvain, Belgium.*

[3] *UK Transplant, Bristol, UK*

[4] *School of Applied Statistics, The University of Reading, Reading, UK*

SUMMARY

Accelerated failure time models with a shared random component are described, and are used to evaluate the effect of explanatory factors and different transplant centres on survival times following kidney transplantation. Different combinations of the distribution of the random effects and baseline hazard function are considered and the fit of such models to the transplant data is critically assessed. A mixture model that combines short-term and long-term components of a hazard function is then developed, which provides a more flexible model for the hazard function. The model can incorporate different explanatory variables

*Correspondence to: Philippe Lambert, Université catholique de Louvain, Institut de Statistique, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve (Belgium). E-mail: lambert@stat.ucl.ac.be Phone: +32-10-47.28.01 Fax: +32-10-47.30.32

and random effects in each component. The model is straightforward to fit using standard statistical software, and is shown to be a good fit to the transplant data.

KEY WORDS: survival analysis; accelerated failure time model; frailty; random effects; Gompertz hazard; transplant survival

1. INTRODUCTION

Accelerated failure time models for time to event data allow a wide range of parametric forms for the hazard function. Random effects, generally known as frailty components, can be introduced into the models, and they can easily be fitted using standard computer software. However, there are situations where certain aspects of the data mean that a hazard function based on a particular probability distribution for the event times is not adequate. In this paper, we propose a mixture model for the hazard function, which allows for different forms of hazard over different periods of time. The model allows for explanatory variables having different effects in each period, and different realisations of random effects can be introduced into the two components of the model.

This work was motivated in part by the analysis of data from UK Transplant on the survival times of a kidney graft following transplantation. This application is therefore used to illustrate the methodology in this paper, but our approach is applicable to other areas of study. The illustration in this paper concerns the outcome of kidney transplants in adults carried out between 1994 and 1996. The response variable of interest is the time from transplant to failure of the graft, this being the earlier of a return to regular dialysis or patient death. Patients were observed for up to 6 years after the transplant.

Because of the large number of centres, 31 in all, and the need to avoid extreme results that might arise from centres which carry out relatively few transplants, centre effects will be regarded as random. A realisation of the random component is shared

by all those who received a transplant in a given centre. This introduces a correlation between the event times for individuals within a centre, and so allows for the transplants being carried out by the same surgical team, for example.

The paper is organised as follows. A description of accelerated failure time models with a shared random component is given in Section 2. The models are then applied to the kidney transplant data and their adequacy critically assessed in Section 3. Section 4 introduces a mixture model, and its use in modelling kidney graft survival times is illustrated in Section 5. Some concluding remarks are presented in Section 6.

2. ACCELERATED FAILURE TIME MODELS WITH FRAILTY

Although parametric proportional hazards models are widely used in medical research, accelerated failure time (AFT) models offer a number of advantages. In particular, they provide for a wider variety of shapes of hazard function than parametric proportional hazards models that assume a particular distribution for survival times, since the family includes distributions with unimodal hazard functions, such as the lognormal distribution. Moreover, the log-linear formulation of such models emphasises that the roles of the regression parameters and dispersion parameters are clearly separated [1]. The regression parameters in an AFT model are also robust towards neglected covariates [2], which is not the case for proportional hazards models [3]. In addition, regression parameters in the proportional hazards model are more sensitive to the distribution of the random component.

Proportional hazards models with frailty effects have been considered by many authors; see, for example, [2, 4]. Some use parametric models for the frailty effect [5], while others describe models with arbitrary distributions [6, 7, 8, 9, 10]. AFT models with frailty effects have received rather less attention. Anderson and Louis [7] use the model in the analysis of bivariate survival data with a parametric or nonparametric frailty distribution, Klein *et al.* [5] consider a survival model based on the lognormal distribution and Pan [11] considers AFT models with gamma frailty effects. Diagnostics for frailty models have received limited attention, but include the work

of Glidden and Self [12] on models with gamma frailty and Kimber and Zhu [13]. A comprehensive review of frailty models is included in Hougaard [14].

2.1. The shared frailty AFT model

The shared frailty AFT model is appropriate in situations where survival times are recorded for groups of individuals who have something in common.

Covariate information may be available at both the subject level and the group level. The vector of values of covariates measured on the jth individual, $j = 1, 2, \ldots, n_i$, in the ith group, $i = 1, 2, \ldots, g$, will be denoted $\boldsymbol{x}_{ij}$, while the vector of values of group-specific covariates will be denoted $\boldsymbol{z}_i$.

In an AFT model, the survivor function at time t, $S(t|\boldsymbol{x}_{ij}, \boldsymbol{z}_i)$, is assumed to be of the form

$$S(t|\boldsymbol{x}_{ij}, \boldsymbol{z}_i) = S_0(t/\psi_{ij}),$$

where $S_0(t)$ is the baseline survivor function associated with reference values of the covariates, and where

$$\psi_{ij} = \psi_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{z}_i) \tag{1}$$

is some function of the covariates. A classical choice for the regression model in equation (1) is

$$\psi_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{z}_i) = \exp(\boldsymbol{\beta}'\boldsymbol{x}_{ij} + \boldsymbol{\gamma}'\boldsymbol{z}_i). \tag{2}$$

The corresponding hazard function is

$$h(t|\boldsymbol{x}_{ij}, \boldsymbol{z}_i) = \frac{1}{\psi_{ij}} h_0(t/\psi_{ij}).$$

This model can equivalently be expressed as a log linear model for the random variable T_{ij}, that is associated with the survival time of the jth individual in the ith group, by writing

$$\log T_{ij} = \mu + \boldsymbol{\beta}'\boldsymbol{x}_{ij} + \boldsymbol{\gamma}'\boldsymbol{z}_i + \sigma\epsilon_{ij},$$

where μ, σ are unknown location and scale parameters, and ϵ_{ij} has a distribution that determines that of T_{ij}. However, anticipating an extension to the model to

Table 1: The parametric forms of the five survival time distributions and the constraint adopted for the three distributions used for the fortitude, α_i.

Distribution	Density function	Constraint
Gamma	$\{e^{\xi\theta}\Gamma(\theta)\}^{-1}t^{\theta-1}\exp\{-t/e^{\xi}\}$	$e^{-\xi}=\theta=\tau$
Inv. Gaussian	$\{\sqrt{(2\pi\theta t^3)}\}^{-1}\exp\{-(t-e^{\xi})^2/(2t\theta e^{2\xi})\}$	$e^{\xi}=1;\ \theta=\tau^{-1}$
Lognormal	$\{\theta t\sqrt{(2\pi)}\}^{-1}\exp\{-(\log t-\xi)^2/(2\theta^2)\}$	$\xi=-\theta^2/2;\ \theta^2=\log(\tau^{-1}+1)$
Log-logistic	$\theta^{-1}\exp\{(\xi-\log t)/\theta\}[1+\exp\{(\xi-\log t)/\theta\}]^{-2}$	
Weibull	$\theta e^{-\xi}\left(te^{-\xi}\right)^{\theta-1}\exp\{-(te^{-\xi})^{\theta}\}$	

be described in Section 4, the formulation based on the hazard function is more convenient.

In the modelling process, it often happens that important influential covariates are missing. We shall assume that these covariates are group-specific and modify the regression equation in (2) to

$$\psi_{ij}(\boldsymbol{x}_{ij},\boldsymbol{z}_i,\alpha_i)=\exp(\eta_{ij})=\exp(\omega_i+\boldsymbol{\beta}'\boldsymbol{x}_{ij}+\boldsymbol{\gamma}'\boldsymbol{z}_i) \tag{3}$$

where $\alpha_i=\exp(\omega_i)$ is distributed across clusters according to some distribution with distribution function $G(\alpha_i)$, and η_{ij} is the linear component of the model. Thus, conditionally on α_i, the AFT model is assumed to hold. The term α_i, which mirrors the role played by a frailty term in the proportional hazards framework, might be termed a *fortitude* since large (small) values of α tend to lead to long (short) lifetimes. We shall refer to our augmented AFT model as a FAFT model for short, where the initial F may stand for frailty or fortitude.

Specific choices can be made for the baseline hazard function and the distribution of the random component, α_i. Constraints may need to be placed on the distribution of α_i to permit identifiability of the parameters; see [15] for details. Specifically, we shall constrain the mean of the distribution of α_i to take the value unity. The parameterisation of the five density functions corresponding to the baseline hazard distributions under consideration are listed in Table 1, together with the constraint used for the three distributions for the α_i to ensure unit mean and variance τ^{-1}.

The model is fitted using the method of maximum likelihood. Let t_{ij} be the observed failure or censoring time for the jth of n_i individuals, $j = 1, 2, \ldots, n_i$, in the ith group, $i = 1, 2, \ldots, g$. Also, let d_{ij} be the corresponding event indicator, so that $d_{ij} = 1$ if a failure is observed and 0 otherwise. Throughout this paper, censoring is assumed to be uninformative. Specifically, conditional on the frailty component, α_i, censoring is assumed to be independent of, and non-informative about, α_i. This ensures that the order of taking account of the censoring, and integrating out the frailty, does not matter [16]. Integrating out the unobserved α_i component, the relevant likelihood function is

$$L = \prod_{i=1}^{g} \int_0^\infty \left\{ \prod_{j=1}^{n_i} h(t_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{z}_i)^{d_{ij}} \; S(t_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{z}_i) \right\} \mathrm{d}G(\alpha_i). \tag{4}$$

Once the maximum likelihood estimates of the parameters have been found, likelihood ratio tests can be used to assess the need for inclusion of a non-degenerate fortitude component, α_i, as well as the other covariates.

The model fitting process can be carried out using packages that include optimisation and numerical integration routines, such as SAS and S-PLUS. The FAFT model can also be fitted using the SAS procedure `proc nlmixed`. To use this procedure, it is only necessary to specify the contribution to the log likelihood function of the jth individual in the ith centre, conditional on α_i. This is assigned to a variable `ll` and a `model` statement is then specified using the option `general(ll)` to describe the distribution of the response variable. A `random` statement is used to define normal random effects, with centres specified using the keyword `subject` in this statement. This will result in a lognormal fortitude component, but other distributions can be fitted using the probability integral transformation. For example, a gamma fortitude component, with unit mean and variance τ^{-1}, is obtained from the SAS statement `g=(1/tau)*gaminv(probnorm(z),tau);`, where `z` is the variable that is declared to have a standard normal distribution in the `random` statement. The integration and optimisation is then carried out by the procedure. SAS code to fit the FAFT model is available from DC.

2.2. Model checking

There are a number of techniques for evaluating the fit of parametric survival models, including analogues of residuals and influence diagnostics [17]. A more informative approach is to stratify the values of the prognostic index, and compare the average survivor functions across the observations in each group with the Kaplan-Meier estimate of the survival times in the groups, as in [18]. This procedure is illustrated later. Another useful method is based on using the values of the linear predictor, or prognostic index, as a covariate in the model as in [19], for example.

In models that include a fortitude component, we first estimate the fortitude associated with each group, so as to be able to use this method for checking model adequacy. We propose to estimate α_i using an empirical Bayes procedure [20], so that the estimate is the mode, $\tilde{\alpha}_i$, of the distribution of α_i, conditional on the data. The linear predictor for the jth individual in the ith group is then

$$\tilde{\eta}_{ij} = \tilde{\omega}_i + \hat{\boldsymbol{\beta}}' \boldsymbol{x}_{ij} + \hat{\boldsymbol{\gamma}}' \boldsymbol{z}_i, \tag{5}$$

where $\tilde{\omega}_i = \log \tilde{\alpha}_i$, and $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}$ are the maximum likelihood estimates obtained from equation (4). Since this procedure is a natural extension of standard methods, it seems reasonable to use it in an informal assessment of model adequacy. However, this might be supplemented by a detailed study of the sensitivity of the diagnostic to departures from the assumed model, so as to explore the effect of introducing the fortitude component. This is beyond the scope of the present paper, and a topic for further work.

3. MODELLING THE SURVIVAL TIMES OF KIDNEY GRAFTS

The principal aims of this analysis are to determine prognostic factors for the transplant survival times (in days) of the first kidney graft in 3511 patients from 31 transplant centres in the UK, and to examine the extent to which the survival times differ between centres. The patients were those aged 18 years or over who received an organ from a cadaveric heartbeating donor.

The data base includes many explanatory variables at the individual level. There are also some at the centre level, but these will not be used in this illustration. Covariates available for each observation include basic demographic information, such as the age and sex of donor and recipient, and variables concerned with the match of donor organ to recipient, such as tissue match. Certain physiological characteristics of donor and recipient were recorded, such as donor cause of death, blood group and recipient diabetes, in addition to the time spent on the waiting list and whether or not the transplant took place at a local centre, in which case the donor organ does not have far to be transported.

The one-year transplant survival rate for this cohort is around 85%, and the event times for 78% of the patients are censored. Patients were observed for at most 2193 days, with censoring generally occurring after 2 (31%), 3 (28%) or 4 (17%) years after surgery. Of the graft failures, 61% failed in the first month, 10% in the second, and 7%, 8%, 6% and 4%, respectively, in each of the four subsequent months.

Parametric models are adopted in the analysis of these data for two reasons. First, an initial analysis of the data showed that the life table estimate of the hazard function is smooth and relatively well behaved, and so a parametric model can be expected to capture this. Second, model checking is so much easier in the context of parametric models, and more reliable [17].

3.1. Fitting FAFT models to the transplant data

Initially, a model selection process was used to identify relevant covariates in an AFT model excluding centres, assuming a Weibull baseline hazard function. The covariates that were found to be relevant were donor age group ($\leqslant 29$, 30–39, 40–49, 50–59, $\geqslant 60$), recipient age group (18–29, 30–39, 40–49, 50–59, $\geqslant 60$), recipient diabetes (no, yes), whether the transplant was carried out at a local centre (no, yes), time on waiting list (<2 years, $\geqslant 2$ years), donor-recipient sex combination (M–M, F–F, F–M, M–F) and tissue match (non-favourable, favourable, highly favourable). Also included on clinical grounds was donor death from a road traffic accident (no, yes),

Table 2: Values of $-2\log\hat{L}$ for the FAFT models considered with relevant covariates.

Survival time	Distribution of random effect			
	None	Gamma	Inv. Gaussian	Lognormal
Weibull	13052.68	13049.73	13049.86	13049.87
Gamma	13059.78	13057.79	13057.85	13057.86
Lognormal	13021.73	13015.44	13015.69	13015.67
Log-logistic	13049.12	13044.82	13045.07	13045.07

although this was not statistically significant. In the modelling process, the first level of each covariate is taken to be the reference category.

Models containing these covariates together with centre effects were then fitted, as were models having different baseline hazards. The FAFT models used in the analysis of the data assume that the exponent of the centre effects in the linear component of the model has either a gamma, inverse Gaussian or lognormal distribution. The values of the statistic $-2\log\hat{L}$, where $\hat{L}$ is the maximised likelihood function for models with the chosen covariates, for a range of survival time distributions, are summarised in Table 2. Of the distributions considered for the survival times, the lognormal appears to be the best choice. We also observe that the value of $-2\log\hat{L}$ differs from that of the Weibull model by more than 30, so that the AFT models are superior to the Weibull model, the only proportional hazards model under consideration. There is a significant random (centre) effect, although there is no great difference in the values of $-2\log\hat{L}$ obtained for the different distributions used to model this effect. This is in line with observations made by various authors; see for example [1]. We therefore propose to use the gamma distribution for the fortitude component, as this leads to the smallest value of $-2\log\hat{L}$.

The set of covariates that were originally identified for the Weibull model were also found to be relevant in this model, although the variable that concerns whether or not the donor was killed in a traffic accident remained non-significant. The parameter estimates and their standard errors for the model with a lognormal baseline survivor function and a gamma fortitude component are given in Table 3.

Table 3: Parameter estimates and standard errors for the selected FAFT model.

Parameter	Estimate	s.e.	p-value
θ (lognormal baseline)	4.583	0.136	–
ξ (lognormal baseline)	11.058	0.463	–
τ (gamma fortitude)	4.046	2.249	–
Donor Age: 30–39	−0.379	0.355	0.284
Donor Age: 40–49	−1.115	0.335	0.001
Donor Age: 50–59	−1.293	0.346	< 0.001
Donor Age: $\geqslant 60$	−1.999	0.386	< 0.001
Recipient Age: 30–39	0.298	0.369	0.418
Recipient Age: 40–49	0.559	0.364	0.126
Recipient Age: 50–59	−0.180	0.349	0.603
Recipient Age: $\geqslant 60$	−1.166	0.368	0.002
Diabetic: Yes	−1.750	0.389	< 0.001
Local: Yes	0.963	0.233	< 0.001
Waiting time: $\geqslant 2$ years	−0.566	0.264	0.032
Road accident: Yes	−0.223	0.334	0.503
Sex combination: F–F	−0.659	0.310	0.034
Sex combination: F–M	−0.488	0.269	0.070
Sex combination: M–F	−0.240	0.304	0.430
Match: highly favourable	1.077	0.251	< 0.001
Match: favourable	0.955	0.510	0.062

In this table, the baseline lognormal hazard function contains the parameters θ, ξ (see Table 1), while the gamma fortitude component has variance τ^{-1}. The remaining estimates should be self-explanatory, but note that the labelling of a sex combination is in the order donor–recipient.

The parameter estimates in Table 3 are actually very similar to those obtained in an AFT model with lognormal baseline hazard but no random centre effect, differing by 0.05 at the most. The standard errors are also practically identical. However, because centre heterogeneity is significant, with a p-value of 0.012 on the basis of a likelihood ratio test, the model that includes centre effects will be used for inference.

If this model passes the checks described in Section 2.2, we would conclude that the risk of failure at a given time increases with the age of donor and that there is a greater risk if the recipient is over 60 years of age. A patient suffering from diabetes or who is on the waiting list for more than two years has a poorer prognosis, but receiving a transplant from a local donor offers a small but significant advantage. There is a greater risk if there is a non-favourable tissue match and if the donor is female.

3.2. Estimation of centre effects and model checking

We first obtain an estimate of the logarithm of the fortitude component for each transplant centre, $\tilde{\omega}_i$ in equation (5). The standard error of $\tilde{\omega}_i$ is then used to obtain a symmetric interval estimate for ω_i in the linear component of the model, and these are used to compare the transplant centres. These estimates, together with their 95% confidence limits, are shown in Figure 1, in which the effects have been arranged in increasing order. Confidentiality restrictions mean that individual centres cannot be identified. Centres with low estimates are those which perform less well, so that the higher the estimate, the greater the fortitude shown by patients in the centre. This figure shows that most of the intervals overlap, and that there are no centres that are performing less well than the others. The heterogeneity between centres therefore arises from the aggregation of small differences between them. Incidentally, although the distribution used for the random centre effects is assumed to be smooth, the

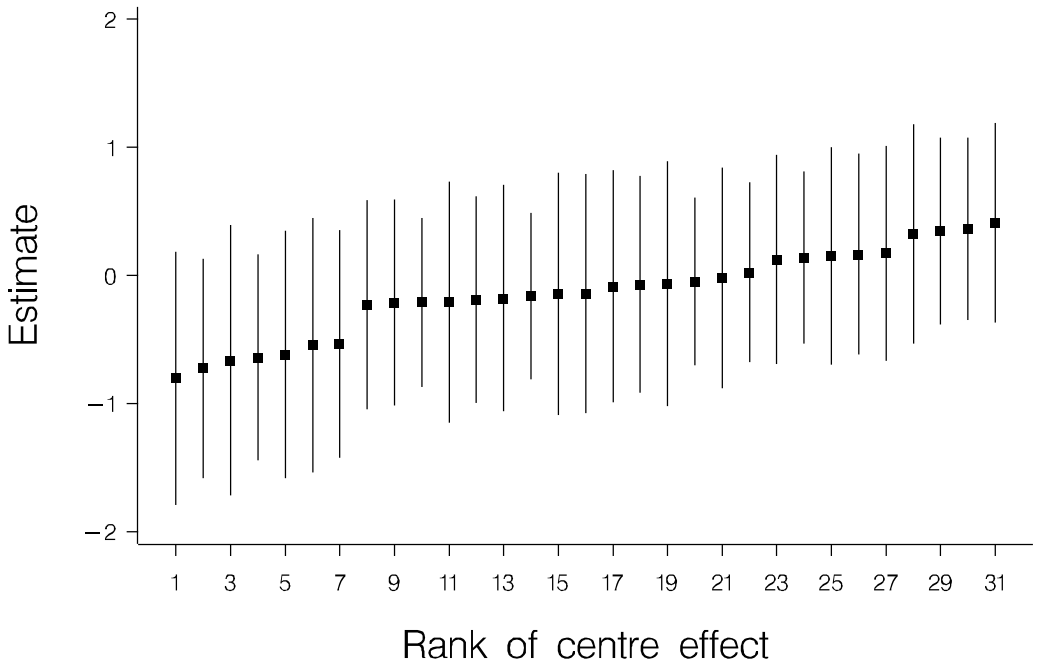

Figure 1: Estimates of centre effects, $\tilde{\omega}_i$, and 95% confidence limits for the 31 transplant centres.

gamma distribution is outlier prone [21, 22]. For this reason, outlying centres would be expected to be shown up in this analysis. The plot also shows some evidence of shrinkage, in that centres with smallest numbers of transplants are the ones with the widest confidence limits.

Using estimates of the fortitude components ω_i, we can obtain values of the linear predictor in equation (5) for each patient. We then stratify these values to give seven categories corresponding to a range of prognoses. The estimated survivor function for the individuals in each stratum are calculated, for times ranging from 0 to 2000 days. Within each stratum, the estimates at each day are averaged to give a fitted survivor function. For clarity, these fitted survivor functions, and the corresponding Kaplan-Meier estimates, for just the first, fourth and seventh strata are shown in Figure 2. Because 71% of the failures took place during the first year, there is more variability in the Kaplan-Meier estimates at times beyond this. Moreover, the form of the fitted survivor functions are largely determined by the failures in the first year.

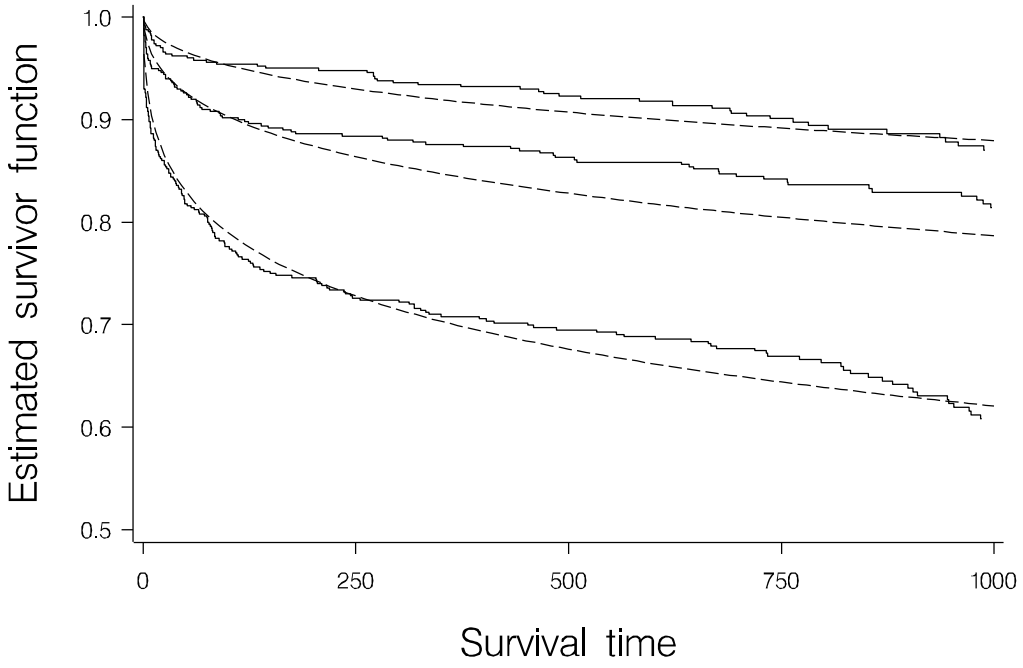

Figure 2: Kaplan-Meier estimates (—) and fitted lognormal survivor functions (– – –) for patients with low, average and good prognosis, respectively. Those with a good prognosis have the highest survival rates.

The FAFT model with a Weibull baseline hazard function and gamma fortitude fails to track the fast declining survivor function in the first 4 months. This explains why the lognormal distribution, with its ability to have a large and quickly decreasing hazard at an early stage, was selected.

4. A MIXTURE MODEL FOR THE HAZARD FUNCTION

Although reasonably satisfactory in its description of the hazard rate for the transplant data during the first few months following transplant, the lognormal distribution is much less convincing at later times. A model that allows the hazard function to decline sharply over the short term, but to then decline more gradually, is likely to be more suitable. Possible ways of modifying the model include adopting a spline model for the baseline hazard function [23], or a semi parametric model [24]. However, since

the covariates that influence the hazard function may not be the same at different times of follow up, a model that incorporates this feature is desirable. Indeed, in the particular context of kidney transplantation, analyses are normally carried out for distinct epochs of follow up in order to account for this [25, 26]. This feature could be modelled through the use of time-varying coefficients in the FAFT model. Although we foresee difficulties in model fitting, and in the selection of an appropriate function of time to use in the modelling process, this is an interesting topic for further research.

For these reasons, we propose a mixture model for the hazard function, and a possible parametric form for such a hazard is

$$h(t|\boldsymbol{x}_{ij}, \boldsymbol{z}_i) = \zeta_{ij}^{-1} \mathrm{e}^{-\lambda t} + \frac{1}{\psi_{ij}} h_0(t/\psi_{ij}) \tag{6}$$

for the jth individual in the ith group. The first term in this mixture model is a Gompertz hazard function which models the initial fast decline in the hazard rate. The parameter λ in this *short-term hazard* determines the rate of decline, while ζ_{ij} is a scale parameter. The second term in (6) models the hazard over a much longer period, and it is convenient to refer to it as the *long-term hazard*. This component is simply an AFT model discussed in previous sections, and parameterised as in Table 1.

The parameter ζ_{ij} may also depend on covariates measured at the centre and group levels. Moreover, the covariates that affect the short term hazard may not be the same as those which are important in the long-term hazard. We may therefore take

$$\zeta_{ij} = \zeta_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{z}_i) = e^{\zeta} \exp(\boldsymbol{\beta}_1' \boldsymbol{x}_{ij} + \boldsymbol{\gamma}_1' \boldsymbol{z}_i), \tag{7}$$

where e^{ζ} is the baseline scale parameter in the Gompertz hazard function, and we now write

$$\psi_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{z}_i) = \exp(\boldsymbol{\beta}_2' \boldsymbol{x}_{ij} + \boldsymbol{\gamma}_2' \boldsymbol{z}_i).$$

In many applications, the random effect might be anticipated to differ in the short and long term. An extension of the model in equation (6) to incorporate random effects in both components allows this to be studied. In the context of the transplant study, these random effects will relate to centre differences, which may not necessarily

be the same in the two components. For example, a centre that deals with more difficult cases, not adequately reflected in the covariates adopted in the model, might appear to be performing less well in the short term, but this may be compensated by an above average performance in the longer term.

Random effects α_{i1} and α_{i2} with respective marginal distributions $G_1(\alpha_{i1})$ and $G_2(\alpha_{i2})$ can be introduced into the two components of the model. This leads to the model in equation (6) with

$$\zeta_{ij} = \zeta_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{z}_i, \alpha_{i1}) = \alpha_{i1} e^{\zeta} \exp(\boldsymbol{\beta}_1' \boldsymbol{x}_{ij} + \boldsymbol{\gamma}_1' \boldsymbol{z}_i) \tag{8}$$

and

$$\psi_{ij} = \psi_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{z}_i, \alpha_{i2}) = \alpha_{i2} \exp(\boldsymbol{\beta}_2' \boldsymbol{x}_{ij} + \boldsymbol{\gamma}_2' \boldsymbol{z}_i). \tag{9}$$

Alternatively, the linear components of the model can be expressed as

$$\omega_{ik} + \boldsymbol{\beta}_k' \boldsymbol{x}_{ij} + \boldsymbol{\gamma}_k' \boldsymbol{z}_i, \quad k = 1, 2,$$

where $\omega_{ik} = \log \alpha_{ik}$.

Adopting a bivariate distribution $G_{12}(\alpha_{i1}, \alpha_{i2})$ for the two random components will enable the extent of any correlation between these two components to be investigated. However, the two components of the mixture model are not orthogonal, and so a correlation between α_{i1} and α_{i2} may well be a result of the structure of the model, rather than the data. Care is therefore needed in interpreting such a correlation.

The corresponding survivor function is

$$S(t|\boldsymbol{x}_{ij}, \boldsymbol{z}_i) = \exp\left\{(\lambda\zeta_{ij})^{-1}\left[e^{-\lambda t} - 1\right]\right\} S_0(t/\psi_{ij}),$$

and the model can be fitted by maximising the likelihood function, integrated out over both α_{i1} and α_{i2}, given by

$$L = \prod_{i=1}^{g} \int_0^\infty \int_0^\infty \left\{ \prod_{j=1}^{n_i} h(t_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{z}_i)^{d_{ij}} \; S(t_{ij}|\boldsymbol{x}_{ij}, \boldsymbol{z}_i) \right\} \mathrm{d}G(\alpha_{i1}) \mathrm{d}G(\alpha_{i2}).$$

As for the FAFT model, this mixture model can be fitted using software that includes facilities for optimisation and numerical integration. The SAS procedure `proc nlmixed` can again be used to fit the mixture model, with the random component declared to have a bivariate normal distribution and then transformed as appropriate.

5. MODELLING THE TRANSPLANT DATA USING THE MIXTURE MODEL

The mixture model proposed in Section 4 is now used in the analysis of the data on kidney graft survival times. A variable selection procedure is adopted in models without centre effects, in order to ascertain which covariates are needed in the two components of the model. For this purpose a lognormal baseline hazard function is used for the long-term hazard component. The variables associated with tissue match are fundamental to organ allocation, and are included in both the short and long term components of the model. Mixture models with the chosen sets of covariates, but with Weibull, Gamma and log-logistic baseline survivor functions for the long-term component are then compared with the lognormal model. Again, it turns out that the lognormal model is superior, with a value of $-2\log\hat{L}$ of 12932.9. Note that the $-2\log\hat{L}$ statistic has a value that is substantially lower than that for the basic FAFT model fitted in Section 3.1 and given in Table 2, even after allowing for the additional parameters that are being fitted through the introduction of the short-term hazard component.

Random effects are then added to both components of the model, with gamma fortitude components used at the outset. A correlation between the two fortitude components is then introduced, so that $(\log\alpha_{i1}, \log\alpha_{i2})$ are modelled using a bivariate normal distribution. The estimated correlation is negative, but not significantly so ($p = 0.51$). There is therefore no evidence that centres who perform less well in the short term make up for it in the longer term, and as mentioned earlier, any correlation could simply be due to the model structure. In further modelling, the short- and long-term fortitude components will be taken to be independent.

Different choices for the distribution of the short-term fortitude component are then considered, in the absence of a long-term fortitude. The gamma distribution is again found to lead to the smallest value of $-2\log\hat{L}$, but there is very little difference between gamma and lognormal fortitudes. When a short-term gamma fortitude is added to the model, the $-2\log\hat{L}$ statistic is reduced from 12932.9 to 12917.2, a change that is highly significant ($p < 0.001$). The significance of this effect is due to

there being many small differences between individual centres, rather than one or two centres being out of line; see the comments in Section 3.2.

Next, a long-term gamma fortitude component is added to the model. This leads to a further reduction in the value of the $-2 \log \hat{L}$ statistic of 0.8, which is not significant. There is therefore no evidence of centre heterogeneity in the longer term.

Parameter estimates and their standard errors for the fitted mixture model with a lognormal model for the baseline survivor function in the long-term hazard, and gamma fortitude in the short-term component only, are given in Table 4.

The p-values in this table, which are confirmed by likelihood ratio tests, indicate that donor age, whether the transplant uses a locally retrieved organ, recipient diabetes and time on waiting list all affect the short term hazard component. In the short term, the condition of the donated organ at transplantation will have a strong influence on transplant survival. Since this tends to be correlated with donor age and the distance that the organ has had to be transported, it is not surprising that donor age and the variable indicating whether or not the transplant uses a locally retreived organ, are particularly significant. The results also show that the effect of the variable 'Local' is not persistent, that the prognosis also depends on the demographic characteristics of the recipient at transplantation, and that degree of tissue match is important. All these variables act in much the same way as in the basic FAFT model described in Section 3.2. While the estimates in Table 4 are interpreted in terms of the effect of the covariates on each component of the hazard, we observe that clinical explanations for these effects are frequently debated in the transplantation literature.

For many, though not all, of the variables in Table 4, the sum of the short and long term parameter estimates are approximately equal to the estimates given in Table 3. From a Taylor expansion of the model in equation (6), it can be seen that this occurs when $h_0(t)$ has an exponential behaviour, which it does in this example; see Figure 3. But in general, there is no reason why the estimates from the two components should sum to the overall value.

To illustrate the hazard functions in the mixture model, Figure 3 shows the overall

Table 4: Parameter estimates and their standard errors for the mixture model with lognormal long-term baseline survivor function, and short-term gamma fortitude.

	Short term hazard			Long term hazard		
Parameter	Estimate	s.e.	p-value	Estimate	s.e.	p-value
λ (Gompertz baseline)	0.208	0.032				
ζ (Gompertz baseline)	4.713	0.307				
θ (lognormal baseline)				3.443	0.183	
ξ (lognormal baseline)				10.701	0.373	
τ (gamma fortitude)	3.817	1.641	0.027			
Donor Age: 30–39	−0.154	0.316	0.630	−0.143	0.309	0.648
Donor Age: 40–49	−0.452	0.278	0.114	−0.564	0.289	0.060
Donor Age: 50–59	−0.313	0.304	0.311	−0.865	0.297	0.007
Donor Age: $\geqslant 60$	−0.772	0.300	0.015	−1.183	0.347	0.002
Recipient Age: 30–39	–	–	–	0.568	0.331	0.096
Recipient Age: 40–49	–	–	–	0.552	0.321	0.096
Recipient Age: 50–59	–	–	–	−0.064	0.306	0.837
Recipient Age: $\geqslant 60$	–	–	–	−1.196	0.316	0.001
Diabetic: Yes	−0.731	0.291	0.018	−1.183	0.378	0.004
Local: Yes	0.795	0.203	0.001	–	–	–
Waiting time: $\geqslant 2$ years	−0.563	0.199	0.008	–	–	–
Sex combination: F–F	–	–	–	−0.562	0.284	0.057
Sex combination: F–M	–	–	–	−0.602	0.235	0.016
Sex combination: M–F	–	–	–	−0.217	0.272	0.431
Match: highly favourable	0.483	0.232	0.046	0.615	0.216	0.008
Match: favourable	−0.069	0.336	0.840	1.003	0.487	0.048

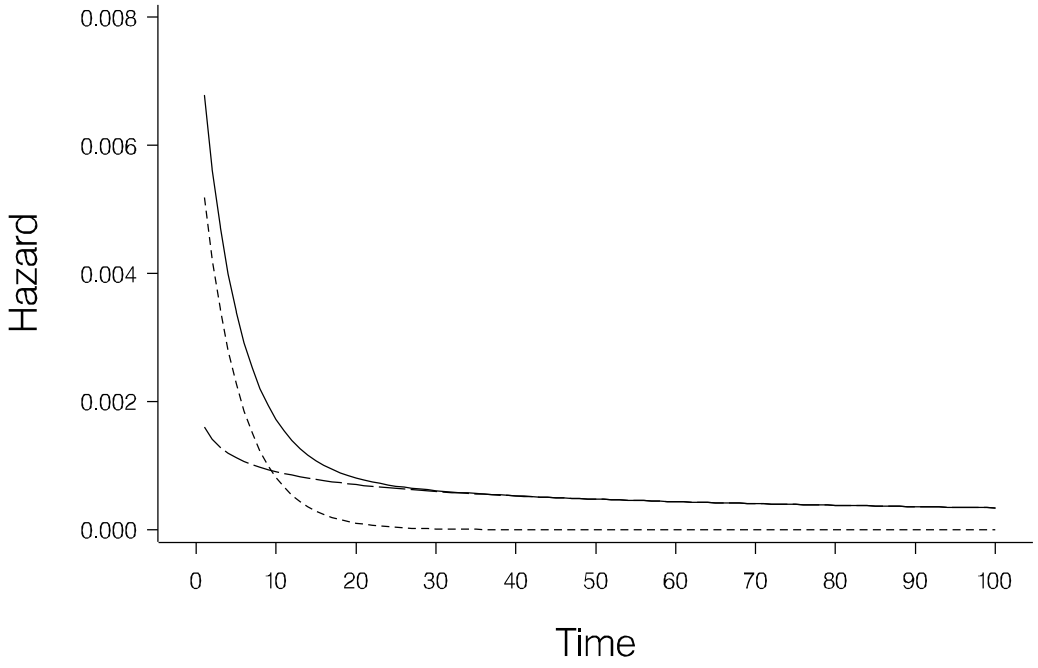

Figure 3: The overall (—), short-term (---) and long-term (– – –) hazard functions for a patient in Centre 1 in the reference category of each covariate for the fitted mixture model.

fitted hazard function, together with the short-term and long-term components for an individual in Centre 1 with covariates set to the reference categories. Note that the time axis in this figure only extends to 100 days, so as to show more clearly the different shapes of the hazard functions over this period. For this particular set of covariate values, the short-term hazard dominates for approximately two weeks, during which time close to 30% of the failures occur. The long term hazard has a persistent effect, and so the covariates in this component have some effect on the hazard function over the whole time period. However, the covariates in the short term component will have a greater impact on the hazard of failure in the period immediately following the transplant.

As before, the adequacy of the fitted model can be examined using the method described in Section 3.2. The resulting graph, shown in Figure 4, confirms that the mixture model is a much better fit over the entire range of the survival times, although

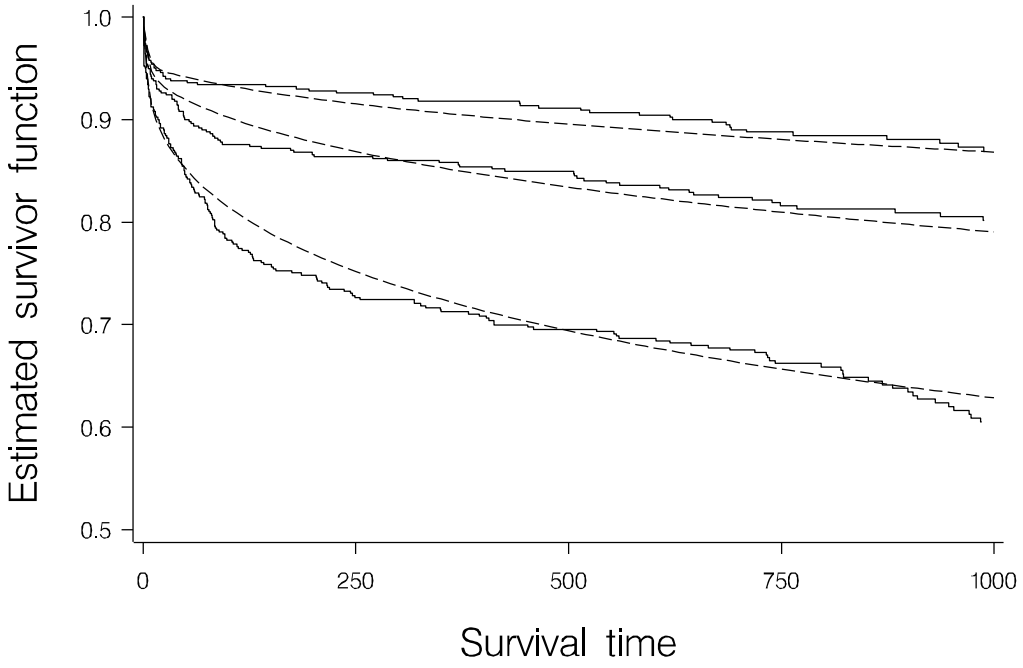

Figure 4: Kaplan-Meier curves and fitted survivor functions for the fitted mixture model.

the fit is not so good between 100 and 250 days. The fit over the first few months is better than that shown in Figure 2, and in the remaining period the fitted survivor functions are considerably closer to the Kaplan-Meier estimates.

Again, an estimate of the fortitude component for each transplant centre can be obtained by computing the posterior mode of the distribution of the random effects in the short-term component of the mixture model, as in Section 2.2. Estimates of the random effects in the linear predictor, $\tilde{\omega}_{i1}$, are plotted together with their associated 95% confidence limits in Figures 5. The centre effects have again been arranged in ascending order of magnitude. This figure shows that there is no one centre that is substantially better or worse than the others. Note that this plot is not directly comparable with that in Figure 1, since Figure 1 is based on an AFT model, while Figure 5 is derived from the Gompertz component of the mixture model. However, it is the differences between centres displayed in a particular figure that are important.

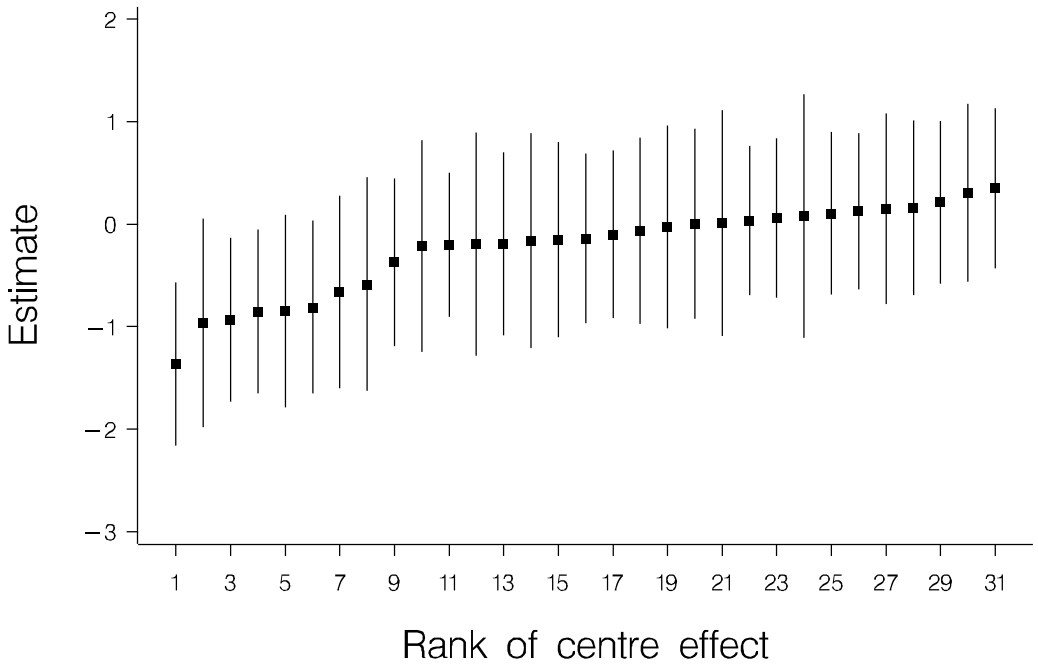

Figure 5: Short-term estimates of centre effects ($\tilde{\omega}_{i1}$) and their 95% confidence limits for the 31 transplant centres.

6. DISCUSSION

In this paper, we have described the role of accelerated failure time models with random effects in the analysis of data on kidney graft survival times. These models provide a flexible approach to the parametric analysis of survival data with frailty. Very often, the choice of distribution for the random component will not be critical, as was seen in our analysis of the transplant data. Models with normally distributed random effects in the linear component may then be adopted, and have the advantage of being straightforward to fit using SAS.

The model checking procedure described in Section 2.2 provides a simple but effective approach to assessing the fit of the models. This procedure may also highlight any weaknesses in the parametric specification of the model, as in our illustration. This approach may be supplemented by plots of Cox-Snell, martingale or deviance residuals [17], but in our experience these are not very sensitive to departures from

an assumed model.

The mixture model proposed in Section 4 allows a much wider variety of shapes of hazard function to be fitted, while maintaining the advantages of fully parametric models. Such models may allow non-standard features of the hazard function to be captured. Moreover, allowing the effects of covariates to differ between the two components can provide new insights into the data. Use of the mixture model has been illustrated in the context of a multi-centre study, and allowing for different centre effects in the two components of the model reveals some interesting features.

A worthwhile extension of the FAFT model would be the development of a non-parametric form for the baseline survivor. Lawless [24] provides an excellent description of a semi-parametric version of the AFT model, but this is not straightforward to fit using existing software packages. However, a parametric choice for the baseline forces the statistician to spend more time on model checking which often yields a better understanding of the data generating process.

ACKNOWLEDGMENTS

We thank UK Transplant, Bristol for making the data set available to us, and Mark Belger and Frances Seeney in particular for helpful discussions. This paper was written while Philippe Lambert was visiting the School of Applied Statistics at the University of Reading with the support of a grant of the Fonds National de la Recherche Scientifique (FNRS, Belgium). The authors are particularly grateful to Sheila Bird and the referees for their very helpful comments and suggestions. We also acknowledge contributions from the Royal College of Surgeons Working Party on Centre Effects during the time when this work was being undertaken.

REFERENCES

1. Keiding, N., Andersen, P. K. and Klein, J. P. 'The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates', *Statistics in Medicine*, **16**, 215–224 (1997).

2. Hougaard, P. 'Fundamentals of survival data', *Biometrics*, **55**, 13–22 (1999).

3. Hougaard, P., Myglegaard, P. and Borch-Johnsen, K. 'Heterogeneity models of disease susceptibility with application to diabetic nephropathy', *Biometrics*, **50**, 1178–1188 (1994).

4. Aalen, O. O. 'Effects of frailty in survival analysis', *Statistical Methods in Medical Research*, **3**, 227–243 (1994).

5. Klein, J. P., Pelz, C. and Zhang, M. 'Modeling random effects for censored data by a multivariate normal regression model', *Biometrics*, **55**, 497–506 (1999).

6. Pickles, A. and Crouchley, R. 'A comparison of frailty models for multivariate survival data', *Statistics in Medicine*, **14**, 1447–1461 (1995).

7. Anderson, J. E. and Louis, T. A. 'Survival analysis using a scale change random effects model', *Journal of the American Statistical Association*, **90**, 669–679 (1995).

8. Sargent, D. J. 'A general framework for random effects survival analysis in the Cox proportional hazards setting', *Biometrics*, **54**, 1486–1497 (1998).

9. Horowitz, J. L. 'Semiparametric estimation of a proportional hazard model with unobserved heterogeneity', *Econometrica*, **67**, 1001–1028 (1999).

10. Walker, S. G. and Mallick, B. K. 'Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing', *Journal of the Royal Statistical Society B*, **59**, 845–860 (1997).

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

12. Glidden, D. V. and Self, S. G. 'Semiparametric estimation in the Clayton-Oakes failure time model', *Scandinavian Journal of Statistics*, **26**, 363–372 (1999).

13. Kimber, A. C. and Zhu, C. Diagnostic for a weibull frailty model, *in* U. Dixit and M. Satam (eds), 'Statistical Inference and Design of Experiments', Narosa Publishing House, New Delhi, India, pp. 36–46, (1999).

14. Hougaard, P. *Analysis of Multivariate Survival Data*, Springer-Verlag, New-York, 2000.

15. Heckman, J. J. and Singer, B. The indentification problem in econometric models for duration data, *in* W. Hildebrand (ed.), 'Advances in Econometrics', Cambridge University Press, pp. 39–77, (1982).

16. Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sorensen, T. I. A. 'A counting process approach to maximum likelihood estimation in frailty models', *Scandinavian Journal of Statistics*, **19**, 25–43 (1992).

17. Collett, D. *Modelling Survival Data in Medical Research, 2nd edn.*, Chapman & Hall/CRC, London, 2003.

18. Copas, J. B. and Heydari, F. 'Estimating the risk of reoffending by using exponential mixture models', *Journal of the Royal Statistical Society, Series A*, **160**, 237–252 (1997).

19. Errington, R. D., Ashby, D., Gore, S. M., Abrams, K. R., Myint, S., Bonnett, D. E., Blake, S. W. and Saxton, T. 'High energy neutron treatment for pelvic cancers: study stopped because of increased mortality', *British Medical Journal*, **302**, 1045–1051 (1991).

20. Carlin, B. P. and Louis, T. *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall/CRC, Boca Raton, 1996.

21. Neyman, J. and Scott, E. L. Outlier proneness of phenomena and of related distribution, *in* J. Rustagi (ed.), 'Optimising Method in Statistics', Academic Press, New York, (1971).

22. Barnett, V. and Lewis, T. *Outliers in Statistical Data, 3rd edn.*, Wiley, Chichester, 1994.

23. Royston, P. and Parmar, M. K. B. 'Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects', *Statistics in Medicine*, **21**, 2175–2197 (2002).

24. Lawless, J. F. *Statistical Models and Methods for Lifetime Data, 2nd edn.*, Chapman & Hall/CRC, Boca Raton, 1996.

25. Gore, S. M., Pocock, S. J. and Kerr, G. R. 'Regression models and non-proportional hazards in the analysis of breast cancer survival', *Applied Statistics*, **33**, 176–195 (1984).

26. Wilks, W. R., Gore, S. M. and Bradley, B. A. 'Renal transplant rejection-transient immunodominance of HLA mismatches', *Transplantation*, **50**, 141–146 (1990).