INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



DISCUSSION PAPER

0223


# EVALUATION OF KIDNEY GRAFT SURVIVAL USING ACCELERATED FAILURE TIME MODELS WITH RANDOM EFFECTS


P. LAMBERT, D. COLLETT, A. KIMBER and R. JOHNSON



http://www.stat.ucl.ac.be

# Evaluation of kidney graft survival using accelerated failure time models with random effects

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

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

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

[3] *UK Transplant, Bristol, UK*

SUMMARY

Accelerated failure time models with a shared random component are used to evaluate the effect of explanatory factors and different transplant centres on the survival times of a kidney graft following 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. Lack of fit of these models leads to the development of a mixture model that combines short-term and long-term components of a hazard function. The model can incorporate different explanatory variables and random effects in each component. The extent of correlation between the random effects in each

*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

component can also be investigated. The model is straightforward to fit using standard 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, which are 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. We therefore 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 by the analysis of data from UK Transplant on the survival times of a kidney graft following transplantation. UK Transplant coordinates organ transplantation in the UK and the Republic of Ireland, and maintains a national data base on transplantation that can be analysed to assist in decision making with regard to organ allocation and patient care. Of particular interest to UK Transplant are factors affecting the survival times of transplanted organs, and the extent of differences in graft survival rates between different transplant centres. This paper focuses on the analysis of data on 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, including death with a functioning graft. In most cases, graft failure does not result in the death of the patient, since the patient can be put on dialysis until a further organ becomes available. Patients were observed for up to 6 years after the transplant. The values of many explanatory variables were recorded, including demographic, biochemical and disease characteristics of both donor and recipient, and variables that describe the match between donor and recipient.

Because of the large number of centres, and the need to avoid extreme results that might arise from centres who 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. Although the centres used are not a random sample of transplant centres, the use of a random effect is justified by arguing that the centre effects are drawn from a random time period.

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 proportional hazards models are widely used in medical research, accelerated failure time (AFT) models offer a number of advantages over them. In particular, they provide for a wider variety of shapes of hazard function than parametric proportional hazards models, and 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].

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 non-parametric 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. For example, the groups may be formed from genetically related subjects as in litters of animals, observations made on paired organs, or repeated event times within an individual, but in the context of the transplant study they refer to the different transplant centres.

Covariate information may be available at both the subject level and the group level. The vector of values of covariates measured on the $j$th individual, $j = 1, 2, \ldots, n_i$, in the $i$th 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

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

In situations where the AFT model is a reasonable choice of model for survival times in a particular application, it often happens that important influential covariates are missing in (2). These might not be considered because they are *a priori* unknown to the investigator or are difficult or expensive to collect. We shall assume that these covariates are group-specific and modify the regression equation 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 $\eta_{ij}$ is the linear component of the model. Thus, conditionally on $\alpha_i$, the AFT model is assumed to hold. The term $\alpha_i$, which mirrors the role played by a frailty term in the proportional hazards framework, might be termed a *fortitude* since large (small) values of $\alpha$ 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.

Note that our specification is different from the proposal of Pan [11]. He introduces a gamma distributed frailty component into an AFT model by assuming proportionality of the conditional hazards associated with different clusters when the observed covariates take identical values. His model is therefore such that

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

where $\psi_{ij}$ is as given in equation (2).

Specific choices can be made for the baseline hazard function and the distribution of the random component, $\alpha_i$. Constraints may need to be placed on the distribution of $\alpha_i$ to permit identifiability of the parameters; see [15] for details. Specifically, we shall constrain the mean of the distribution of $\alpha_i$ 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 $\alpha_i$ to ensure unit mean and variance $\tau^{-1}$.

The model is fitted using the method of maximum likelihood. Let $t_{ij}$ be the observed failure or censoring time for the $j$th of $n_i$ individuals, $j = 1, 2, \ldots, n_i$, in the

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

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

$i$th 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. Integrating out the unobserved $\alpha_i$ 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, $\alpha_i$, 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 $i$th individual in the $j$th centre, conditional on $\alpha_i$. 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 $\tau^{-1}$, 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

A range of possible distributions for the baseline survival times and the fortitude component have been summarised. The combination of baseline and fortitude distributions with the largest maximised likelihood could be selected to make inferences about covariate effects. However, it may be that none of the FAFT models considered is an adequate fit, and so it is important to check the adequacy of the selected model.

When no heterogeneity or fortitude is detected, the traditional techniques for model checking can be used. When the retained covariates are categorical and not too numerous, the fitted survival distributions can be compared to the corresponding Kaplan-Meier estimate for each covariate combination. When there are many or continuous covariates, the linear predictor can be stratified to correspond to, say, a poor, average and good prognosis. That summary covariate can then be used as the sole explanatory variable in an AFT model with the selected baseline survival distribution. The fitted survival distribution and the Kaplan-Meier estimate can be compared for each value of the covariate associated with prognosis.

In models that include a fortitude component, we must 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 $\alpha_i$ using an empirical Bayes procedure [16], so that the estimate is the mode of the distribution of $\alpha_i$, conditional on the data. More precisely, suppose that

$$L(\alpha_i, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}} | \boldsymbol{t}_i)$$

denotes the likelihood contribution of individuals in group $i$, for a given value of the fortitude $\alpha_i$, with regression parameters fixed at their marginal maximum likelihood estimates, that is their maximum likelihood estimates after integrating out the frailty component. If the fitted fortitude distribution with variance $\tau^{-1}$ and density function

$g(\alpha_i|\hat{\tau})$ is treated as the 'prior', then the 'posterior' density of $\alpha_i$ is proportional to

$$L(\alpha_i, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}|\boldsymbol{t}_i)\, g(\alpha_i|\hat{\tau}).$$

We shall use the mode, $\tilde{\alpha}_i$, of this posterior density as an estimate of $\alpha_i$. The linear predictor for the $i$th individual in the $j$th group is then

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

where $\tilde{\omega}_i = \log \tilde{\alpha}_i$. These methods are now used in modelling the data on kidney graft survival times.

## 3. MODELLING THE SURVIVAL TIMES OF KIDNEY GRAFTS

The principal aims of this analysis are to determine prognostic factors for the 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, but none at the centre level. These covariates 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 antigen 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 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. Figure 1 shows the proportions of failures at event times in the first 6 months of observation.

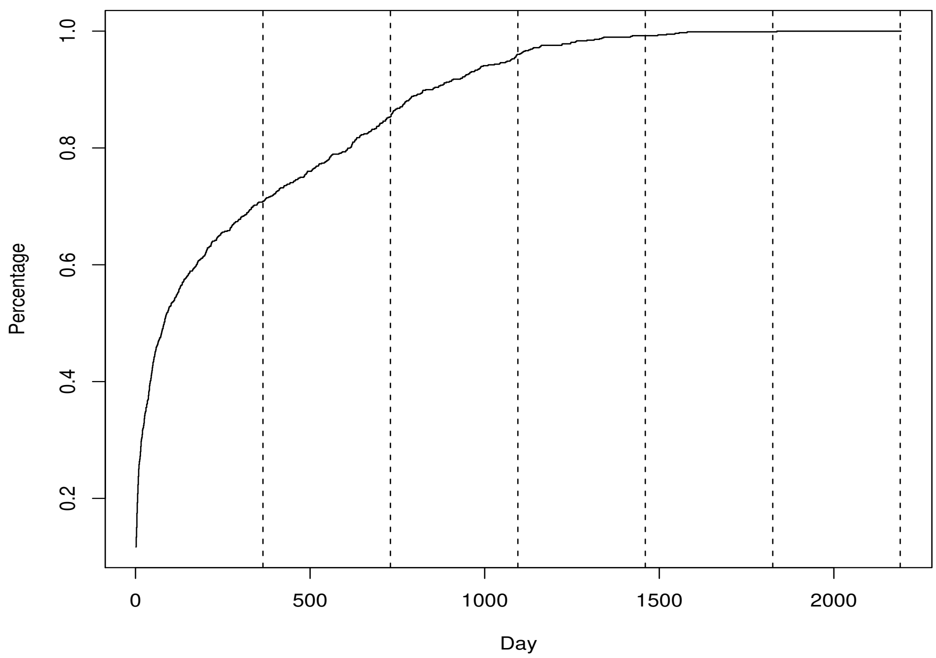


Figure 1: Proportion of patients experiencing graft failure out of the 779 failures observed during the follow-up period.

### 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 death from a road traffic accident (no, yes), 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

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 |

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. 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.

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

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

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

| Parameter | Estimate | s.e. | $p$-value |
|---|---|---|---|
| $\theta$ (lognormal baseline) | 4.583 | 0.136 | – |
| $\xi$ (lognormal baseline) | 11.058 | 0.463 | – |
| $\tau$ (gamma fortitude) | 4.046 | 2.249 | 0.072 |
| 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 |

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 significant advantage. There is a greater risk if there is a non-favourable tissue match and if the donor is female, particularly if the recipient is also 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 $\omega_i$ 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 2, 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.

Using estimates of the fortitude components $\omega_i$, we can obtain values of the linear predictor in equation (5) for each patient. We then stratify these values to give categories corresponding to poor, average and good prognosis, where each stratum contains approximately one third of the patients in the data base. The lognormal AFT model was again fitted to the survival times, with the prognosis category taken as a factor with three levels. A plot of the Kaplan-Meier estimate and of the corresponding fitted lognormal survivor function for the three strata can be seen in Figure 3. The left hand figure shows the plot for the whole observation period whereas the right hand figure gives the plot for the first year only. 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

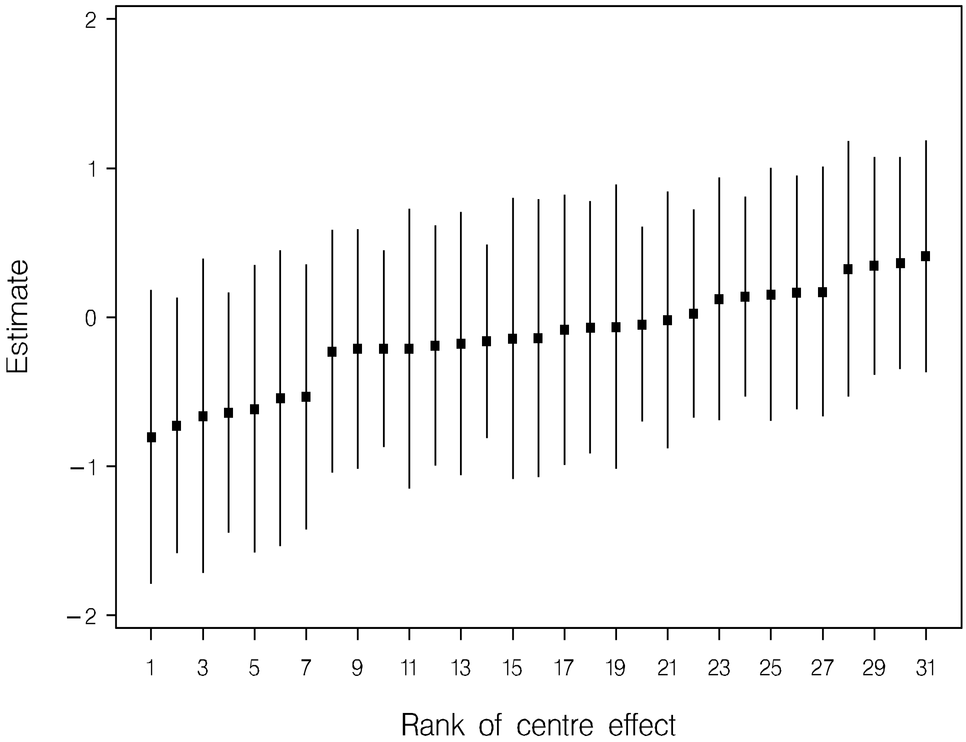


Figure 2: Estimates of centre effects, $\tilde{\omega}_i$, and 95% confidence limits for the 31 transplant centres.

determined by the failures in the first year.

Comparing these figures with those corresponding to other baseline hazard distributions indicates why the lognormal model is superior. For example, consider the FAFT model with a Weibull baseline hazard function and gamma fortitude. After fitting this model to the data, stratifying the individuals and then fitting Weibull models to the resulting three strata gives Figure 4.

This figure clearly shows that the fitted Weibull model 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.

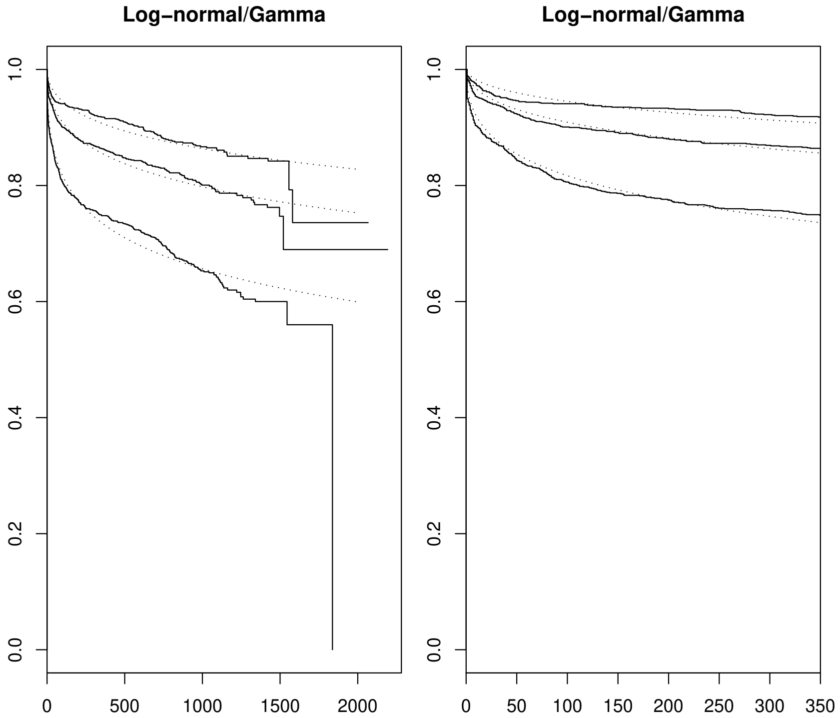


Figure 3: Kaplan-Meier estimates (—) and fitted lognormal survivor functions ($\cdots$) for patients with low, average and good prognosis, respectively. Those with a good prognosis have the greatest survival rates.

## 4. A MIXTURE MODEL FOR THE HAZARD FUNCTION

Although satisfactory in its description of the hazard rate for the transplant data during the first few months following transplant, the lognormal distribution seems 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.

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 $j$th individual in the $i$th group. The first term in this mixture model is a Gompertz hazard function which models the initial fast decline in the hazard rate.

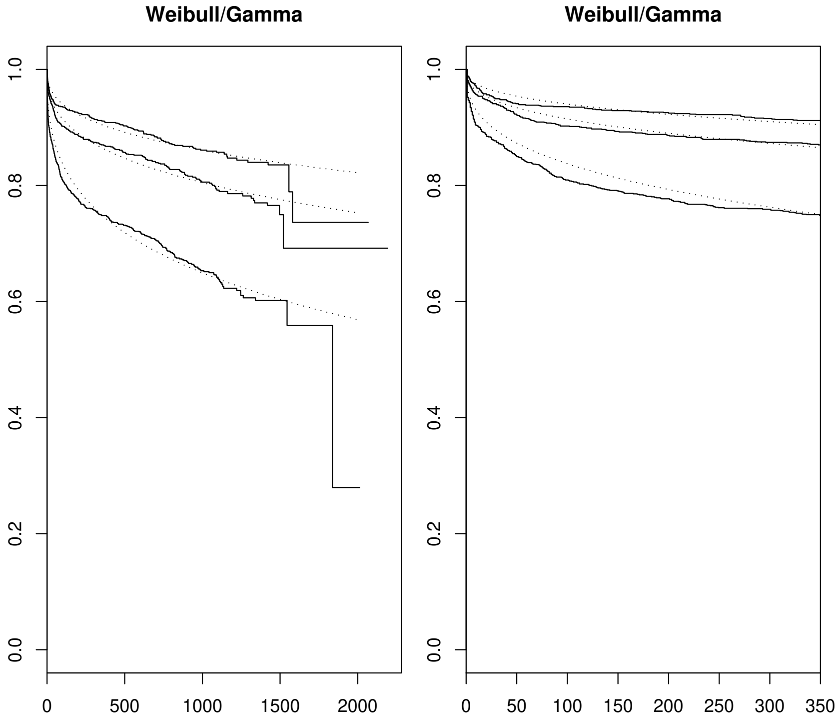


Figure 4: Kaplan-Meier estimates (—) and fitted lognormal survivor functions ($\cdots$) for patients with low, average and good prognosis, respectively. Those with a good prognosis have the greatest survival rates.

The parameter $\lambda$ in this *short-term hazard* determines the rate of decline, while $\zeta_{ij}$ is a scale parameter. The second term in (6) models the *long-term hazard* using the AFT model discussed in previous sections, and parameterised as in Table 1.

The parameter $\zeta_{ij}$ 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 $e^{\zeta}$ is the baseline scale parameter in the Gompertz hazard function, and we now write

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

Again random effects can be included in both components of this model. In the context of the transplant study, these will relate to centre effects, 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 long term.

Random effects $\alpha_{i1}$ and $\alpha_{i2}$ with respective marginal distributions $G_1(\alpha_{i1})$ and $G_2(\alpha_{i2})$ can be introduced into the two components of the model. Moreover, adopting a bivariate distribution $G_{12}(\alpha_{i1}, \alpha_{i2})$ will enable the extent of any correlation between these two components to be investigated. 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 $\omega_{ik} = \log \alpha_{ik}$. 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 $\alpha_{i1}$ and $\alpha_{i2}$, 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. 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 12937.4. 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 ($p = 0.42$). There is therefore only a very slight suggestion that centres who perform less well in the short term make up for it in the longer term. In further modelling, the short- and long-term fortitude components will be taken to be independent.

Next, different choices for the distribution of the short-term fortitude component are 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 12937.4 to 12922.6, a change that is highly significant ($p < 0.001$).

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 in the short term donor age, whether the transplant is local, recipient diabetes and time on waiting list are all important. In the longer term, both the age of the donor and that of the recipient are important, and so are recipient diabetes, the donor-recipient sex combination, and the level of antigen match. These variables act in much the same way as in the basic FAFT model described in Section 3.2.

To illustrate the hazard functions in the mixture model, Figure 5 shows the overall 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. The short-term hazard dominates for approximately two weeks, during which time close to 30% of the failures occur.

As before, the adequacy of the fitted model can be examined by first stratifying the values of the linear predictor in both components of the model. The Kaplan-Meier estimates of the survivor functions for the individuals in these two strata can then be compared with those under the mixture model that contains the covariate associated with the stratification in each component. The resulting graphs are shown in Figure 6, with just two strata being used to make these figures easier to interpret. The left hand graphs correspond to the complete time period, and those on the right hand side to the first year only. The top two graphs refer to the short-term component and bottom graphs to the long-term component. These figures confirm that the mixture model is a much better fit over the entire range of the survival times.

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 7. The centre effects have again been arranged in

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 |
| $\lambda$ (Gompertz baseline) | 0.208 | 0.033 | | | | |
| $\zeta$ (Gompertz baseline) | 4.926 | 0.292 | | | | |
| $\theta$ (lognormal baseline) | | | | 3.456 | 0.183 | |
| $\xi$ (lognormal baseline) | | | | 10.697 | 0.374 | |
| $\tau$ (gamma fortitude) | 3.892 | 1.713 | 0.030 | | | |
| Donor Age: 30–39 | −0.133 | 0.320 | 0.681 | −0.141 | 0.311 | 0.654 |
| Donor Age: 40–49 | −0.470 | 0.278 | 0.101 | −0.539 | 0.289 | 0.072 |
| Donor Age: 50–59 | −0.268 | 0.314 | 0.401 | −0.895 | 0.301 | 0.006 |
| Donor Age: $\geqslant 60$ | −0.788 | 0.305 | 0.015 | −1.185 | 0.349 | 0.002 |
| Recipient Age: 30–39 | – | – | – | 0.548 | 0.332 | 0.108 |
| Recipient Age: 40–49 | – | – | – | 0.529 | 0.321 | 0.109 |
| Recipient Age: 50–59 | – | – | – | −0.087 | 0.307 | 0.778 |
| Recipient Age: $\geqslant 60$ | – | – | – | −1.206 | 0.318 | 0.001 |
| Diabetic: Yes | −0.765 | 0.287 | 0.012 | −1.147 | 0.379 | 0.005 |
| Local: Yes | 0.673 | 0.191 | 0.001 | – | – | – |
| Waiting time: $\geqslant 2$ years | −0.588 | 0.200 | 0.006 | – | – | – |
| Sex combination: F–F | – | – | – | −0.558 | 0.285 | 0.060 |
| Sex combination: F–M | – | – | – | −0.611 | 0.236 | 0.015 |
| Sex combination: M–F | – | – | – | −0.243 | 0.273 | 0.382 |
| Match: highly favourable | – | – | – | 0.733 | 0.213 | 0.002 |
| Match: favourable | – | – | – | 0.957 | 0.480 | 0.055 |

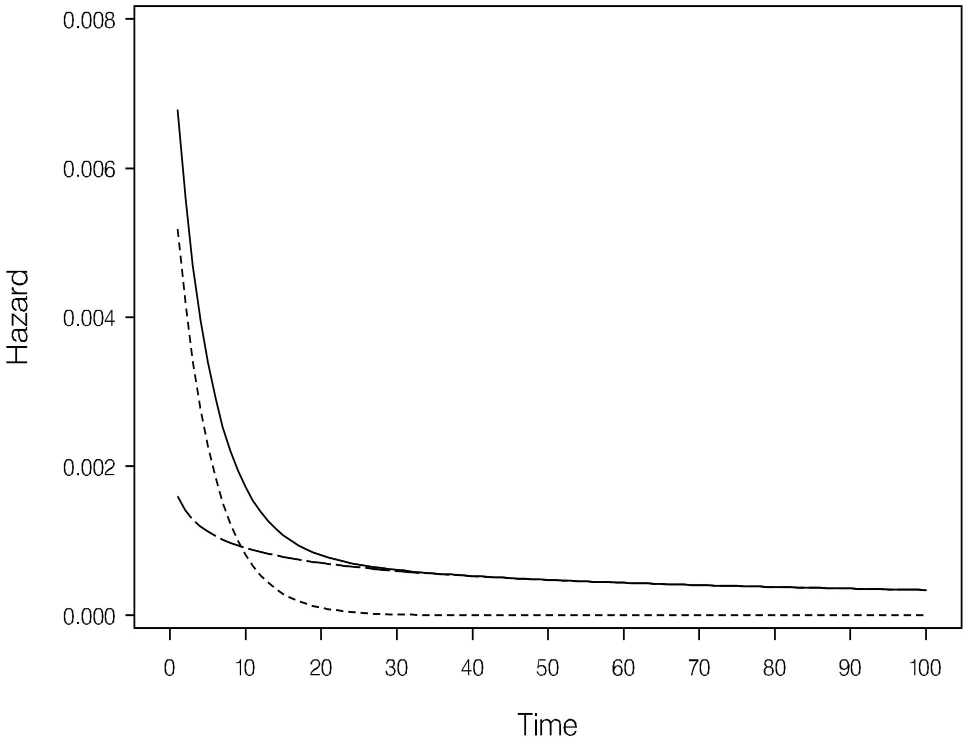


Figure 5: The overall (—), short-term (- - -) and long-term (· · ·) hazard functions for a patient in Centre 1 in the reference category of each covariate for the fitted mixture model.

ascending order of magnitude. This figure shows that there is no one centre that is substantially better or worse than the others.

## 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.

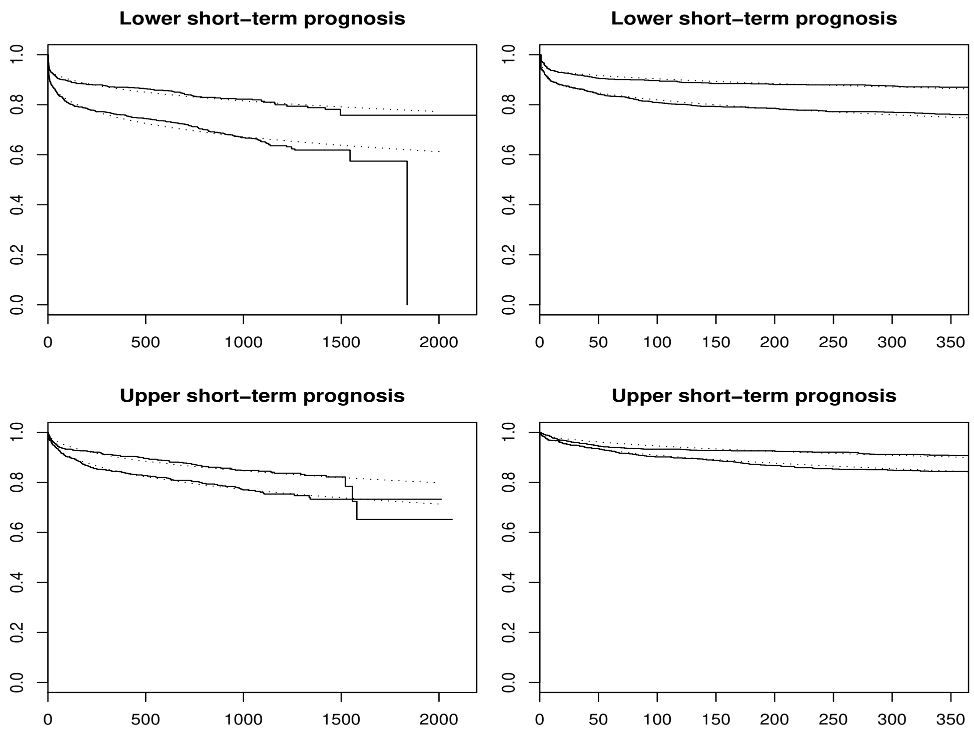


Figure 6: Kaplan-Meier curves and fitted survivor functions corresponding to the two components of the fitted mixture model.

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

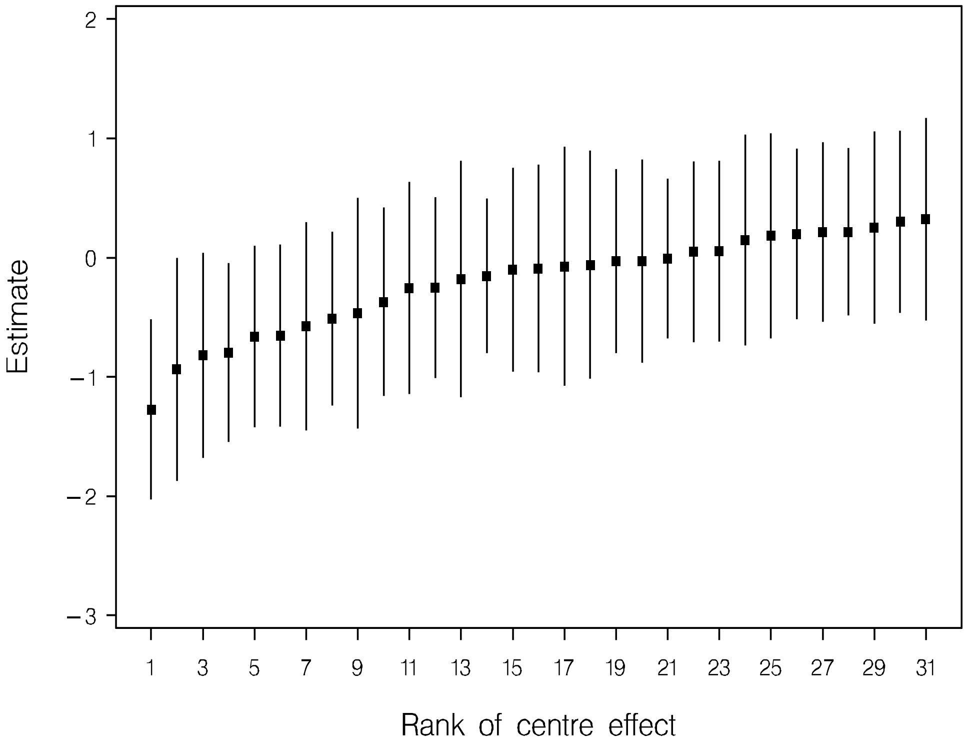


Figure 7: Short-term estimates of centre effects ($\tilde{\omega}_{i1}$) and their 95% confidence limits for the 31 transplant centres.

effects in the two components of the model reveals some interesting features of the data.

A worthwhile extension of the FAFT model would be the development of a non-parametric form for the baseline survivor. 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).

Financial support from the contract 'Projet d'Actions de Recherche Concertes' nr 98/03–217 from the Belgian government, and from the IAP research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

## REFERENCES

1. Keiding, N., Andersen, P. and Klein, J. '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., 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. and Louis, T. 'Survival analysis using a scale change random effects model', *Journal of the American Statistical Association*, **90**, 669–679 (1995).

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

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

10. Walker, S. and Mallick, B. '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. and Self, S. 'Semiparametric estimation in the Clayton-Oakes failure time model', *Scandinavian Journal of Statistics*, **26**, 363–372 (1999).

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

17. Collett, D. *Modelling Survival Data in Medical Research*, Chapman & Hall/CRC, London, 1994.