Lifetime Data Analysis, 2, 241–259 (1996)

Modelling Conditional Distributions in Bivariate Survival

ROBIN HENDERSON henderr1@lancaster.ac.uk
Department of Mathematics and Statistics, University of Lancaster, LA1 4YF, UK

Received November 9, 1995; Revised March 18, 1996; Accepted March 26, 1996

Abstract. Conditional distributions for bivariate survival can be obtained via a model for the joint distribution, or, as has sometimes been suggested, by modelling the conditioned variable directly, with the conditioning variable included as a covariate. A quantitative comparison of estimated covariate effects and predictive distributions under the two approaches is given. The results are illustrated in a novel frailty application.

Keywords: Bivariate Survival, Conditional Distribution, Frailty.

1. Introduction

There is considerable interest in methods for analysing bivariate, or more generally multivariate, survival data. A variety of applications are described in recent work by, for instance, Aalen et al. (1995), Anderson and Louis (1995), Hougaard (1995), Hougaard et al.(1992), McGilchrist (1993), Pickles and Crouchley (1994), Shih and Louis (1995) and Wassell et al. (1995). One area which still needs further work however concerns modelling the conditional distribution of one variable given the other(s), such as the age at death of a daughter given that of the mother, or the time to a second infection given that to a first. Here one may derive the conditional distribution from a model for the joint distribution, or as an alternative could model the conditional distribution directly, incorporating the given failure time as a fixed covariate. The latter technique is perhaps less natural but will usually be simpler to implement, requiring only standard software, and was suggested by discussants of Clayton and Cuzick (1985) with further remarks by Hougaard (1987). The most important question of course is whether the two approaches lead to substantially different inferences in practice: if not then the simpler technique could be recommended.

The aim of this paper is to make a quantitative comparison of the two modelling approaches. We assume a bivariate frailty model to describe association between two survival times and we are interested in the conditional distribution of one variable given an uncensored observation of the other. We compare the exact conditional distribution with the asymptotic fit of a misspecified parametric marginal model for the first variable incorporating the second variable as a covariate.

The original motivation for the work came from the analysis of some data on the survival of 272 lung cancer patients. At diagnosis of disease a subjective survival time prediction P was made by the consulting physician, and this is to be compared with actual survival T. Thus (T, P) can be considered to form bivariate survival data with perhaps very different marginals. One could not argue that T is determined by P, simply that the two are associated,

and perhaps both are linked to some unmeasurable state or "health" of the individual, which leads to the suggestion of a frailty model. Since P is uncensored and available at diagnosis there is a strong case for building information on P, ie indirect information on the frailty, into the predictive distribution of T given other covariates. Hence the interest in conditional distributions.

In Section 2 we compare exact conditional distributions from the most widely used bivariate survival model with the asymptotic fit of a univariate model when there is no censoring. In Section 3 the effect of censoring is investigated, and the different approaches are illustrated using the lung cancer data in Section 4. Some closing remarks in Section 5 complete the paper.

2. Comparison of Conditional Distributions: No Censoring

2.1. Notation and assumptions

Although the results of the following two sections are general, we allow the motivating example to determine the terminology. Thus instead of referring to bivariate data as usual as (T_1, T_2) we will use (T, P), with T being outcome and P being prediction. Note that the conditioning variable, in our case P, in most applications would not be considered to be a "prediction" of the other variable.

Let $\underline{x}$ denote a vector of covariates, including an intercept term, and let $\tilde{\underline{x}}$ be the vector formed by removing the intercept. We assume that the covariates are scaled to zero mean. We use $S(.)$ to denote survivor functions and suppress the dependence on $\underline{x}$ throughout. In agreement with Aalen (1994) we adopt a parametric approach and assume the most widely applied bivariate survival model, namely a gamma frailty variable acting multiplicatively on otherwise independent Weibull hazards (Aalen, 1988, 1994, Aalen et al., 1995, Clayton, 1995, Hougaard et al., 1994, and Shih and Louis 1995). Thus the model is

$$S_{T|Z}(t \mid z) = \exp\{-zt^{\sigma} e^{\underline{\beta}_t^T \underline{x}}\},$$

$$S_{P|Z}(p \mid z) = exp\{-zp^{\delta} e^{\underline{\beta}_p^T \underline{x}}\}$$

and

$$Z \sim \Gamma(\lambda, \lambda).$$

It is then straightforward to integrate out Z (Hougaard, 1987) and show that the joint survivor function is

$$S(t,p) = \frac{\lambda^{\lambda}}{(\lambda + p^{\delta} e^{\underline{\beta}_p^T \underline{x}} + t^{\sigma} e^{\underline{\beta}_t^T \underline{x}})^{\lambda}},$$

that the marginals are

$$S_T(t) = \frac{\lambda^\lambda}{(\lambda + t^\sigma e^{\underline{\beta}_t^T \underline{x}})^\lambda} \text{ and } S_P(p) = \frac{\lambda^\lambda}{(\lambda + p^\delta e^{\underline{\beta}_p^T \underline{x}})^\lambda},$$

and that the conditional survivor function of T given $P = p$ (uncensored) is

$$S_{T|P}(t \mid p) = \frac{(\lambda + p^\delta e^{\underline{\beta}_p^T \underline{x}})^{\lambda+1}}{(\lambda + p^\delta e^{\underline{\beta}_p^T \underline{x}} + t^\sigma e^{\underline{\beta}_t^T \underline{x}})^{\lambda+1}}.$$

We shall refer to this conditional distribution as the *frailty-conditional.*

2.2. *Misspecified model and estimating equations*

Now suppose that the model is misspecified, with P considered to be a covariate in a parametric marginal model for T. The most likely false approach involves fitting a model from the proportional hazards family, and hence we assume a Weibull distribution is selected. Since the distribution of P is likely to be heavily tailed its logarithm is used as covariate. Thus the falsely assumed model is

$$S^F_{T|P}(t \mid p) = \exp(-e^{\underline{\beta}^T \underline{x}} p^\alpha t^\nu),$$

which will be referred to as the *Weibull-conditional.* The parameters $\underline{\beta}$, α and ν are to be estimated from a sample of observation times $t_1, t_2, ..., t_n$, which for the moment are assumed to be uncensored. Differentiating the assumed Weibull log-likelihood with respect to $\underline{\beta}$, α and ν respectively leads to the estimating equations

$$\sum_{i=1}^n \{\underline{x}_i - \underline{x}_i e^{\underline{\beta}^T \underline{x}_i} p_i^\alpha t_i^\nu\} = \underline{0},$$

$$\sum_{i=1}^n \{\log(p_i) - e^{\underline{\beta}^T \underline{x}_i} \log(p_i) p_i^\alpha t_i^\nu\} = 0$$

and

$$\sum_{i=1}^n \{\frac{1}{\nu} + \log(t_i) - e^{\underline{\beta}^T \underline{x}_i} \log(t_i) p_i^\alpha t_i^\nu\} = 0.$$

Hence the maximum likelihood solutions $\hat{\underline{\beta}}$, $\hat{\alpha}$ and $\hat{\nu}$ are asymptotically consistent to the solutions $\underline{\beta}_0$,α_0 and ν_0 to

$$E[\underline{X}] = E[\underline{X} e^{\underline{\beta}^T \underline{X}} P^\alpha T^\nu] \tag{1}$$

$$E[\log(P)] = E[e^{\underline{\beta}^T \underline{X}} \log(P) P^{\alpha} T^{\nu}] \tag{2}$$

and

$$E[\frac{1}{\nu} + \log(T)] = E[e^{\underline{\beta}^T \underline{X}} \log(T) P^{\alpha} T^{\nu}], \tag{3}$$

where expectation is with respect to the joint distribution of $\underline{X}$,P and T.

2.3. *Solution of the equations*

We now derive the solutions to (1)–(3) when the frailty model of Section 2.1 is in fact correct. We begin by conditioning on $\underline{X}$, though it will become apparent that this conditioning is unnecessary as most important features are independent of the distribution of $\underline{X}$. The following results can be derived via standard integrals (Gradshteyn and Ryzhik 1994, 3.194.3 and 4.253.6) and are valid provided both ν/σ and α/δ are greater than -1, and also $\lambda - \nu/\sigma - \alpha/\delta$ is positive:

$$E[P^{\alpha} T^{\nu} \mid \underline{X}] = \lambda^{\frac{\nu}{\sigma}+\frac{\alpha}{\delta}+1} e^{-\frac{\nu}{\sigma}\underline{\beta}_t^T \underline{X} - \frac{\alpha}{\delta}\underline{\beta}_p^T \underline{X}} \left\{ \frac{\Gamma(\frac{\nu}{\sigma}+1)\Gamma(\frac{\alpha}{\delta}+1)\Gamma(\lambda - \frac{\nu}{\sigma} - \frac{\alpha}{\delta})}{\Gamma(\lambda+1)} \right\}, \tag{4}$$

$$E[\log(P) \mid \underline{X}] = \frac{1}{\delta}\{\log(\lambda) - \underline{\beta}_p^T \underline{X} - C - \psi(\lambda)\}, \tag{5}$$

$$E[\log(T) \mid \underline{X}] = \frac{1}{\sigma}\{\log(\lambda) - \underline{\beta}_t^T \underline{X} - C - \psi(\lambda)\}, \tag{6}$$

$$E[\log(P) P^{\alpha} T^{\nu} \mid \underline{X}] = \frac{1}{\delta}\{\log(\lambda) - \underline{\beta}_p^T \underline{X} + \psi(\frac{\alpha}{\delta}+1) - \psi(\lambda - \frac{\nu}{\sigma} - \frac{\alpha}{\delta})\} E[P^{\alpha} T^{\nu} \mid \underline{X}]; \tag{7}$$

$$E[\log(T) P^{\alpha} T^{\nu} \mid \underline{X}] = \frac{1}{\sigma}\{\log(\lambda) - \underline{\beta}_t^T \underline{X} + \psi(\frac{\nu}{\sigma}+1) - \psi(\lambda - \frac{\nu}{\sigma} - \frac{\alpha}{\delta})\} E[P^{\alpha} T^{\nu} \mid \underline{X}] \tag{8}$$

where $\Gamma(.)$ is the usual gamma function, C is Euler's constant and $\psi(.)$ is the digamma function, namely $\psi(\lambda) = d\log(\Gamma((\lambda))/d\lambda$.

Solution of equations (1)–(3) can now be separated into two stages: we can derive an equation for ν_0 which can be solved numerically, and can then obtain explicit solutions for α_0 and $\underline{\beta}_0$. First recall that (1) defines a system of equations: one for each element of $\underline{X}$. Since an intercept term is included it follows from the first of these that α_0, ν_0 and $\underline{\beta}_0$ must satisfy

$$1 = E[e^{\underline{\beta}_0^T \underline{X}} P^{\alpha_0} T^{\nu_0} \mid \underline{X}].$$

Hence we can replace the right–most conditional expectations in (7) and (8), ie $E[P^{\alpha} T^{\nu} \mid \underline{X}]$, with $\exp(-\underline{\beta}_0^T \underline{X})$. Substitution of (5)–(8) into (2) and (3) then gives, after some cancellation,

$$C + \psi(\lambda) + \psi(\tfrac{\alpha_0}{\delta} + 1) - \psi(\lambda - \tfrac{\nu_0}{\sigma} - \tfrac{\alpha_0}{\delta}) = 0 \tag{9}$$

and

$$\frac{\sigma}{\nu_0} + \psi(\tfrac{\alpha_0}{\delta} + 1) - \psi(\tfrac{\nu_0}{\sigma} + 1) = 0. \tag{10}$$

Since these equations do not depend upon the value of $\underline{X}$ we can remove the conditioning and see that the solutions α_0 and ν_0 do not depend on the distribution of $\underline{X}$, nor the values of $\underline{\beta}_t$ or $\underline{\beta}_p$, and moreover that the true shape parameters σ and δ are involved only through the ratios ν_0/σ and α_0/δ. Further simplification is also possible by recalling that, for any $u \neq 0$, $\psi(u+1) = \psi(u) + 1/u$ (Gradshteyn and Ryzhik 1994, 8.365). Applying this to $\psi(\frac{\nu_0}{\sigma} + 1)$ in (10) it follows that

$$\psi(\tfrac{\alpha_0}{\delta} + 1) - \psi(\tfrac{\nu_0}{\sigma}) = 0,$$

from which we see $\alpha_0/\delta = \nu_0/\sigma - 1$. Substitution in (9) then gives

$$C + \psi(\lambda) + \psi(\tfrac{\nu_0}{\sigma}) - \psi(\lambda + 1 - 2\tfrac{\nu_0}{\sigma}) = 0. \tag{11}$$

This has to be solved numerically, which is straightforward since Fortran subroutines to evaluate $\psi(.)$ and its derivatives are available in the NAG library and so an iterative procedure is easy to implement.

Having obtained α_0 and ν_0 we can obtain $\underline{\beta}_0$ explicitly. The conditional expectation (4) can be substituted in (1) to give

$$\begin{aligned} E_X[\underline{X}] &= \lambda^{\frac{\nu_0}{\sigma} + \frac{\alpha_0}{\delta} + 1} \left\{ \frac{\Gamma(\frac{\nu_0}{\sigma} + 1)\Gamma(\frac{\alpha_0}{\delta} + 1)\Gamma(\lambda - \frac{\nu_0}{\sigma} - \frac{\alpha_0}{\delta})}{\Gamma(\lambda + 1)} \right\} E_X[\underline{X} e^{\underline{\gamma}^T \underline{X}}] \\ &= A E_X[\underline{X} e^{\underline{\gamma}^T \underline{X}}] \end{aligned}$$

say, where $\underline{\gamma} = \underline{\beta}_0 - \frac{\nu_0}{\sigma}\underline{\beta}_t - \frac{\alpha_0}{\delta}\underline{\beta}_p$. Recall that the covariates are scaled to have zero mean, and so on excluding the intercept term we must have $E[\tilde{\underline{X}}] = \underline{0}$, whence $\underline{\gamma}^T \underline{X}$ must be independent of $\underline{X}$, which implies $\tilde{\underline{\gamma}} = \underline{0}$ and so

$$\tilde{\underline{\beta}}_0 = \tfrac{\nu_0}{\sigma}\tilde{\underline{\beta}}_t + \tfrac{\alpha_0}{\delta}\tilde{\underline{\beta}}_p. \tag{12}$$

The intercept itself follows by writing $\underline{\gamma}^T \underline{X} = \gamma_1$ from which $\gamma_1 = -\log(A)$ and so

$$\left(\underline{\beta}_0\right)_1 = \tfrac{\nu_0}{\sigma}\left(\underline{\beta}_t\right)_1 + \tfrac{\alpha_0}{\delta}\left(\underline{\beta}_p\right)_1 - \log(A). \tag{13}$$

Note that given $\underline{\beta}_t$ and $\underline{\beta}_p$ the estimated covariate coefficients are again independent of the distribution of $\underline{X}$. Hence we have obtained the asymptotic values of the parameter estimates under the misspecified model: ν_0 solves (11),α_0/δ is just $\nu_0/\sigma - 1$ and $\underline{\beta}_0$ is given by (12) and (13).

Table 1. Comparison of Survival Curves

	Frailty-conditional			Weibull-conditional			
	$\theta_p = .1$	.5	.9	$\theta_p = .1$	.5	.9	
$\lambda = 100.0$	0.899	0.900	0.901	0.898	0.899	0.900	$.9 = \theta_t$)
	0.497	0.499	0.504	0.497	0.501	0.504	.5
	0.098	0.099	0.103	0.098	0.100	0.102	.1
4.0	0.880	0.895	0.928	0.858	0.888	0.906	.9
	0.429	0.478	0.603	0.431	0.521	0.581	.5
	0.060	0.081	0.163	0.056	0.107	0.156	.1
1.0	0.826	0.898	0.978	0.748	0.870	0.935	.9
	0.277	0.444	0.826	0.285	0.547	0.748	.5
	0.012	0.033	0.277	0.004	0.073	0.285	.1
0.5	0.770	0.918	0.996	0.671	0.873	0.969	.9
	0.157	0.432	0.957	0.176	0.553	0.874	.5
	0.001	0.008	0.356	0.000	0.012	0.362	.1

2.4. *Comparison of true and fitted conditional distributions*

Consider the conditional survivor function of T given P at some covariate value $\underline{x}_0$. Let $t^* = t^*(\theta_t)$ be a fixed percentile of the marginal distribution of T, i.e. the solution to

$$S_T(t^*) = \frac{\lambda^\lambda}{(\lambda + (t^*)^\sigma e^{\underline{\beta}_t^T \underline{x}})^\lambda} = 1 - \theta_t,$$

and define $p^* = p^*(\theta_p)$ similarly. Then it is convenient to compare the actual frailty-conditional and fitted Weibull-conditional survivor curves at these percentiles, ie to compare $S(t^* \mid p^*)$ with $S^F(t^* \mid p^*)$. This is because the comparison is independent of the choice of $\underline{\beta}_t$, $\underline{\beta}_p$, σ, δ, $\underline{x}_0$ and the distribution of $\underline{X}$, depending only on λ, θ_t and θ_p.

Table 1 gives comparative survival values for various choices of λ, θ_t and θ_p. Several points are worth noting.

1. When λ is large the variance of the frailty term is small and there is only low dependence of T on P. Thus true and fitted conditional survivor functions are close to the marginal survivor function of T, as expected.

2. As λ decreases the dependence becomes stronger and low P is associated with low T and so on.

3. For λ not large the fitted survivor curve tends to underestimate the true function at early values of t^* and overestimate later. The crossover is relatively late for large p^*.

4. The difference between $S(t^* \mid p^*)$ and $S^F(t^* \mid p^*)$ can be quite marked when both t^* and p^* are near the median values in particular.

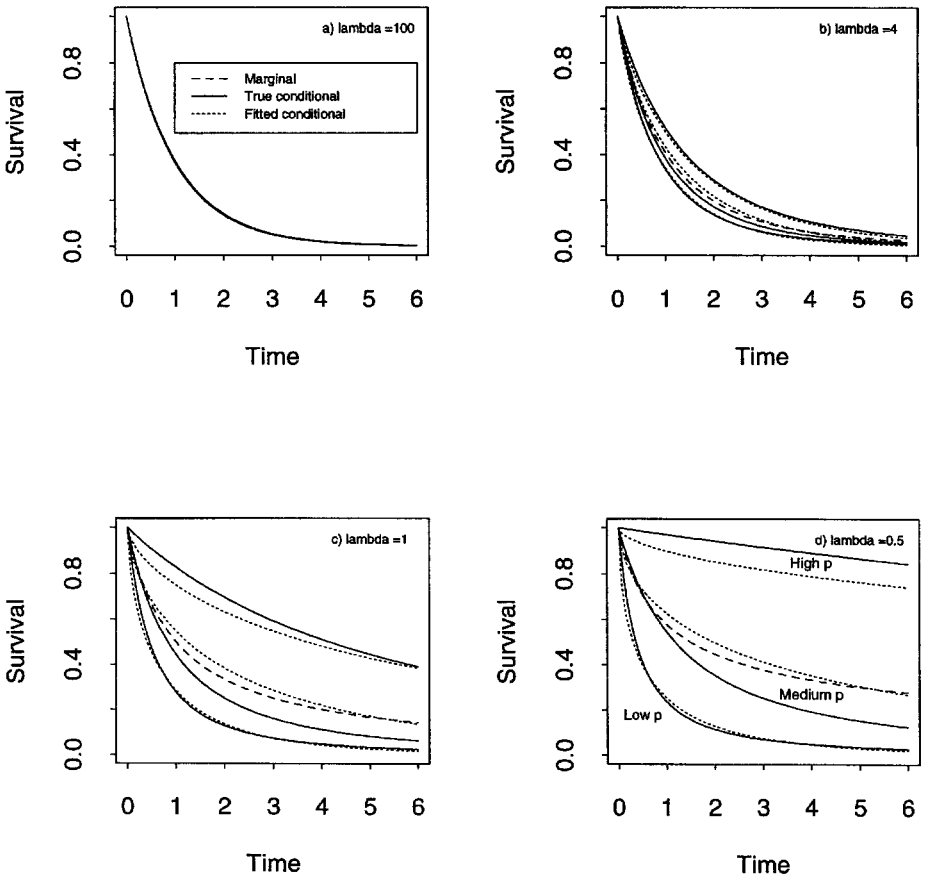


Figure 1. True and fitted survival functions.

Figure 1 gives further information, showing true (solid line) and fitted (dotted) conditional survivor functions together with the marginal survivor function of T (dashed). Values of λ of 100,4,1 and 0.5 are illustrated in the four plots, each of which shows three pairs of conditional distributions, corresponding to low p ($\theta_p = 0.1$), medium p ($\theta_p = 0.5$) and high p ($\theta_p = 0.9$). Conclusions are as summarised above. Simulation results, not given, indicate that the asymptotic fit is very close to the average fit in even small samples.

Finally, for reference Figure 2 shows how the relative shape parameters ν_0/σ and α_0/δ change as λ increases. The plot shows both ν_0/σ and α_0/δ, marked "Weibull", against the frailty standard deviation, i.e. $1/\sqrt{\lambda}$.(The plot marked "Cox" will be discussed later). When the standard deviation is zero there is no dependence between T and P and so $\nu_0 = \sigma$

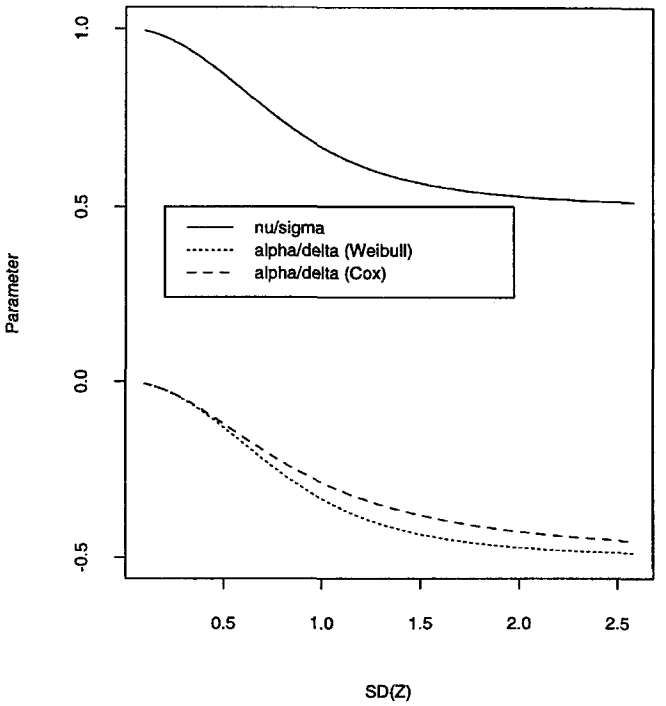


Figure 2. Shape parameters of fitted distribution.

and $\alpha_0 = 0$ as required. As the standard deviation increases we find ν_0 tending to 0.5σ and α_0 tending to -0.5δ, which can be confirmed formally through series expansion of the $\psi(.)$ functions. Thus $\underline{\beta}_0$ approaches $0.5(\underline{\beta}_t - \underline{\beta}_p)$ for strong association between T and P, showing that the apparent effect of any covariate under the Weibull-conditional depends on the difference between its effects on T and P.

3. Comparison of Conditional Distributions: Censoring Allowed

3.1. Assumptions

Now suppose that observed survival is complicated by the presence of independent random censorship. Letting Δ be the censoring indicator, the estimating equations become

$$E[\Delta \underline{X}] = E[\underline{X} e^{\underline{\beta}^T \underline{X}} P^{\alpha} S^{\nu}],$$

$$E[\Delta \log(P)] = E[e^{\underline{\beta}^T \underline{X}} \log(P) P^{\alpha} S^{\nu}],$$

and

$$E[\frac{\Delta}{\nu} + \Delta \log(S)] = E[e^{\underline{\beta}^T \underline{X}} \log(S) P^{\alpha} S^{\nu}],$$

where $S = \min(T, U)$, with U being the censoring time. Some of these expectations can be evaluated for certain parametric models for U. For instance, if we assume U has a Weibull distribution with the same shape parameter σ as actual failure time T, i.e.

$$S_U(u) = \exp(-\theta u^{\sigma}),$$

then it can be shown that

$$E[\Delta \mid \underline{X}] = \lambda \mu^{\lambda} e^{\mu} \Gamma(-\lambda, \mu)$$

where $\mu = \theta \lambda e^{\underline{\beta}_t^T \underline{X}}$ and $\Gamma(.,.)$ is the incomplete gamma function. Thus, for simulation purposes, θ can be selected to give any required proportion of censoring. Most of the other integrations are intractable however under this or any other reasonable model for U, and so we use simulation to investigate the effect of censoring.

The results below assume no covariate, $\beta_t = \beta_p = 0$, $\sigma = \delta = 1$, and exponential censorship. Qualitatively similar results are obtained for other scenarios. Fitted curves are obtained from samples of 200,000 observations.

3.2. *Moments*

The mean and variance of the true frailty conditional distribution of T given $P = p$ are

$$E[T \mid P = p] = \frac{(\lambda + p)}{\lambda}$$

and

$$Var(T \mid P = p) = \begin{cases} \frac{(\lambda+p)^2(\lambda+1)}{\lambda^2(\lambda-1)} & \lambda > 1 \\ \infty & 0 < \lambda \leq 1 \end{cases}$$

whilst the mean and variance of the fitted Weibull conditional distribution are

$$E^F[T \mid P = p] = \Gamma(1 + \frac{1}{\nu_0})/(e^{\beta_0} p^{\alpha_0})^{1/\nu_0}$$

and

$$Var^F(T \mid P = p) = \{\Gamma(1 + \frac{2}{\nu_0}) - \Gamma(1 + \frac{1}{\nu_0})^2\}/(e^{\beta_0} p^{\alpha_0})^{2/\nu_0}.$$

Table 2. Moments of Conditional Distributions

		Mean			Variance		
		$\theta_p = .1$	.5	.9	$\theta_p = .1$	.5	.9
$\lambda = 100.0$	FC	1.00	1.01	1.02	1.02	1.03	1.07
	WC 0%	1.00	1.01	1.02	1.01	1.04	1.05
	WC 50%	1.00	1.01	1.02	1.00	1.03	1.05
$\lambda = 4.0$	FC	1.03	1.19	1.78	1.76	2.36	5.27
	WC 0%	0.99	1.33	1.64	1.30	2.33	3.56
	WC 50%	0.94	1.21	1.47	0.99	1.67	2.44
$\lambda = 1.0$	FC	1.11	2.00	10.00	∞	∞	∞
	WC 0%	0.94	2.83	8.49	2.13	19.19	172.69
	WC 50%	0.88	1.88	4.02	1.09	4.99	22.89
$\lambda = 0.5$	FC	1.24	4.00	100.00	∞	∞	∞
	WC 0%	0.91	5.91	76.48	2.84	118.83	19931.74
	WC 50%	0.89	3.05	16.61	1.32	15.65	465.21

Table 2 presents some numerical results to compare true moments with fitted, with and without censoring. Three values of p are considered, corresponding to 10%, 50% and 90% quantiles of the marginal distribution of P for each λ. Rows labelled FC refer to the true frailty-conditional, and WC 0% and WC 50% to the Weibull-conditional with those degrees of random censoring. Although only 0% and 50% censoring are tabulated, other values lead to the same conclusions, which are:

1. Obviously, as P increases both the true mean and variance increase, as the conditional distribution shifts to the right.

2. Both conditional means and variances increase rapidly as λ decreases and the dependence between T and P increases.

3. Without censoring the fitted Weibull-conditional underestimates both the mean and variance at the extreme values of P. For P more central however the fitted mean can overestimate the true value.

4. As censoring is introduced the fitted mean and variance decrease. The distribution is shifted to the left and the long right-hand tail brought in.

3.3. Fitted distributions

Figure 3 shows the effect of censoring more clearly. Here $\lambda = 0.5$ and again we condition on P at the 10%, 50% and 90% quantiles. For low P ($\theta_p = 0.1$) there is little real difference between the true frailty conditional and the fitted Weibull conditional with any amount of censoring. As P increases however the distributions become more heavily tailed and there can be considerable discrepancies. For medium P ($\theta_p = 0.5$) the fitted distribution without

censoring is more heavily tailed than the true distribution. Introduction of censoring brings in the tail of the fitted distribution and thus improves the fit. At high P ($\theta_p = 0.9$) on the other hand the fitted Weibull conditional is already too short tailed without censoring, which when present amplifies the true to fitted difference.

4. Illustration

4.1. Background

The lung cancer data referred to in Section 1 will be used to illustrate the two different approaches in practice. There is now no question of one model being correct and the other not, since the true distribution is unknown. We are interested instead in seeing what differences, if any, there are in practice between fitted conditional distributions or estimated covariate effects.

Outcome T is time in months from diagnosis until death, and there are 272 observations, of which 45 were censored. Predictions P were obtained from one of five experienced physicians. There is no attempt here to compare physicians or to include anything other than independent linear effects of the six covariates used: age, sex (0=F,1=M), activity score (0–4, high score showing poor activity) and presence (1) or absence (0) of anorexia, hoarseness or metastases. Further details are given in Henderson (1995), where measurements of prediction accuracy are discussed, and Henderson and Jones (1995), where outcome is compared with both subjective and model-based predictions.

4.2. Fitted distributions

Table 3 shows the results of fitting the bivariate frailty model to the data. The mle of the estimated frailty parameter λ is 1.253 and a likelihood ratio test of no association is decisively rejected. Considering the covariates, age apparently has little effect on either T or P, but the others have strong influence on at least one of these variables. The physicians were little influenced by the sex of the patient whereas the males in fact survived longer than the females in this group, but they, the physicians, seemed to consider activity score to be a better survival predictor than it turned out to be. Similarly they were a little too influenced by metastases rather than hoarseness, though by and large there was good agreement between the association between covariates and outcome and that between covariates and predictions.

Fitted survival curves are compared with Kaplan-Meier plots in Figure 4. The data were split into three roughly equally sized groups according to prognostic indicator $PI_i = \hat{\underline{\beta}}^T \underline{x}_i$, i.e. the 90 with lowest risk, 91 of medium risk and 91 of highest risk. Figure 4a shows Kaplan-Meier plots for T using $\hat{\underline{\beta}}_t$ and Figure 4b refers to P, using $\hat{\underline{\beta}}_p$. In each case marginal survival plots from the fitted joint distribution are also shown, assuming average covariate values for each group. The fit is reasonable for both T and P.

Table 4 shows the results of fitting Weibull distributions to T with and without (log)P included as a covariate. Without P the Weibull fit to T gives similar results to the frailty

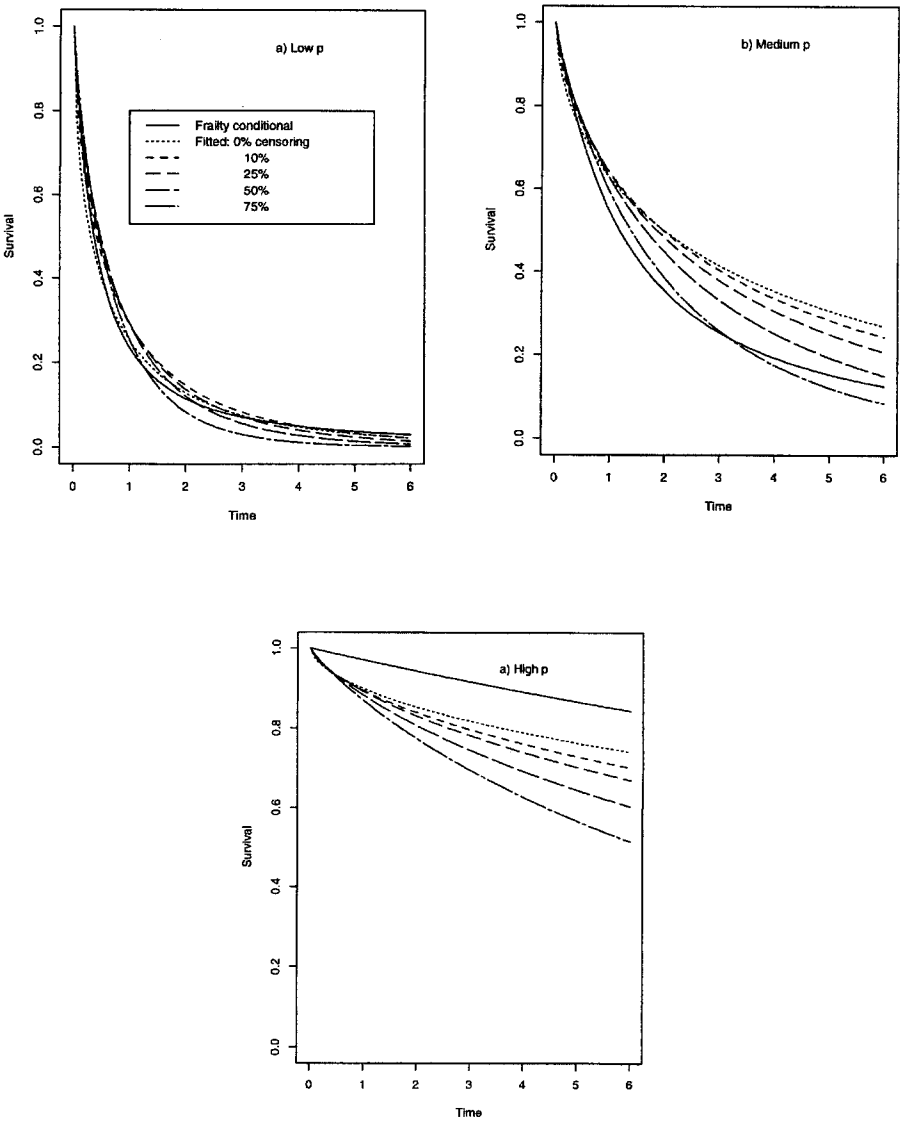
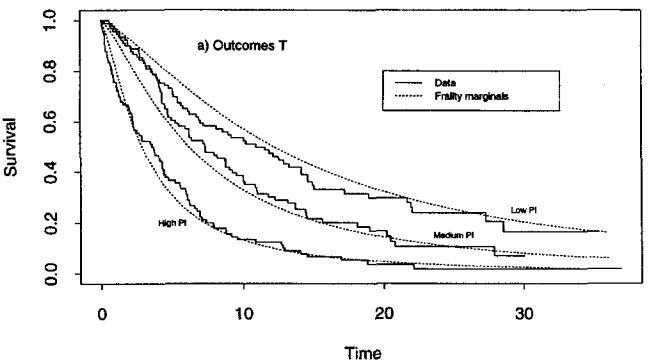


Figure 3. Effect of censoring on Conditional survival.

Table 3. Parameter Estimates for Bivariate Frailty Model

	Outcome T		Prediction P	
	Estimate	t-stat	Estimate	t-stat
Constant	-4.222	-5.52	-6.455	-8.10
Age	0.011	1.03	0.004	0.35
Sex	-0.766	-3.03	-0.202	-0.80
Activity	0.566	4.85	0.941	7.80
Anorexia	0.658	3.08	0.775	3.58
Hoarseness	0.814	2.59	0.556	1.81
Metastases	0.407	1.26	0.851	2.72
	$\hat{\sigma} = 1.383$ (se=0.082)		$\hat{\delta} = 2.289$ (se=0.138)	
	$\hat{\lambda} = 1.253$ (se=0.212)		log-likelihood = -659.10	



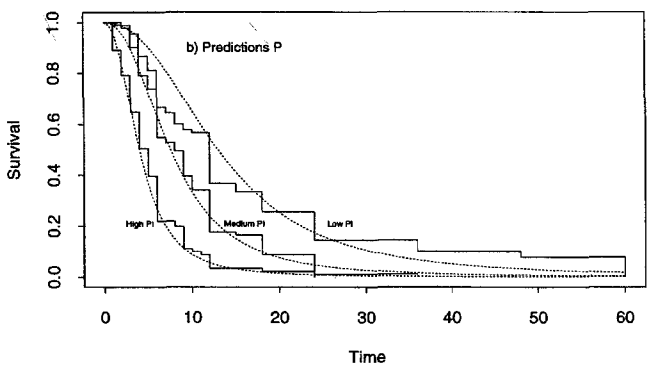


Figure 4. Frailty marginals compared with data.

Table 4. Parameter Estimates for Weibull Models

	Without P		With P		Formula 12
	Estimate	t-stat	Estimate	t-stat	
Constant	-3.401	-5.95	-1.767	-2.87	-
Age	0.009	1.06	0.007	0.86	0.008
Sex	-0.718	-4.01	-0.625	-3.49	-0.549
Activity	0.349	4.27	0.156	1.81	0.172
Anorexia	0.305	2.10	0.198	1.38	0.294
Hoarseness	0.705	3.37	0.561	2.71	0.483
Metastases	0.423	1.92	0.083	0.37	0.072
Log(Prediction)	-	-	-0.676	-7.00	-
	$\hat{\nu} = 1.023$ (se=0.054)		$\hat{\nu} = 1.099$ (se=0.057)		
	log-likelihood = -711.61		log-likelihood= -688.77		

model, except for moderately reduced effects of activity score and anorexia. With P included the activity, anorexia, hoarseness and metastases variables all have less importance, presumably because of their correlation with P, which has highly significant effect. The maximised log-likelihood rises substantially when P is included, indicating that there is indeed information in P which is not already available through the other covariates. For reference, if the other covariates are omitted and only P and a constant term allowed, the estimated effect of $\log(P)$ becomes more significant (coefficient -0.785, t-stat -9.17) but the maximised log-likelihood falls to -703.30. Thus as anticipated there is also information in the covariates which is not available through P alone.

The final column of Table 4 shows covariate coefficients obtained by formula (12), i.e. asymptotic values if the frailty model is assumed to be true with parameters as in Table 3. There is good agreement between these and the estimates actually obtained from the Weibull fit incorporating P.

Figure 5a, obtained by the same method as Figure 4, shows that the Weibull fit incorporating P leads to similar survival curves at average covariates (and P) as the frailty marginals: both fit equally well. A Weibull fit without P gives similar curves to those shown in Figure 5a, but it is interesting to note that a Weibull model for the predictions themselves fits poorly as a result of the long tails (Figure 5b), especially for the low risk group.

A referee has suggested that it might be instructive to compare estimated variances of $\log(T)$ since for a Weibull distribution with shape parameter ϕ this is $\psi'(1)/\phi^2$, independent of covariates. For the bivariate frailty distribution the equivalent result is $\{\psi'(1)+\psi'(\lambda)\}/\sigma^2$ unconditionally and $\{\psi'(1)-\psi'(\lambda+1)\}/\sigma^2$ conditional upon P. Numerical estimates are presented in Table 5. There is a slight gain in estimated precision for the Weibull model when $\log(P)$ is included as a covariate, a more substantial improvement through conditioning when the bivariate model is assumed.

4.3. Conditional distributions

A cross-validatory technique was adopted in comparing conditional distributions, with for each case the parameters being estimated from the rest of the data only, so that the

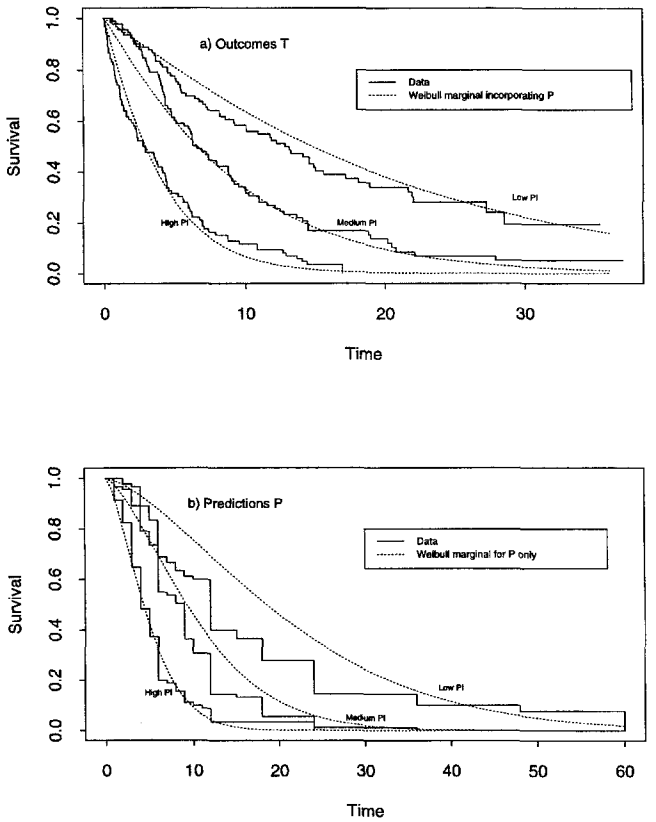


Figure 5. Weibull marginals compared with data.

Table 5. Estimated Variance of $\log(T)$

	Weibull	Frailty
Unconditional	1.57	1.48
Conditional (given P)	1.36	0.57

Table 6. Proportion of Observed S_i and S_i^F in Nominal 25% Intervals.

	0–0.25	0.25–0.5	0.5–0.75	0.75–1
Frailty-conditional S_i	0.274	0.247	0.233	0.244
Weibull-conditional S_i^F	0.249	0.261	0.287	0.203

Table 7. Mean Survival Given P.

		t =3	6	12	24	36
$P \leq 3$	$\bar{S}_{T\|P}(t \mid p)$	0.40	0.19	0.06	0.01	0.00
(n=45)	$\bar{S}^F_{T\|P}(t \mid p)$	0.40	0.18	0.04	0.00	0.00
$3 < P < 24$	$\bar{S}_{T\|P}(t \mid p)$	0.72	0.50	0.26	0.10	0.05
(n=194)	$\bar{S}^F_{T\|P}(t \mid p)$	0.73	0.53	0.29	0.10	0.04
$P \geq 24$	$\bar{S}_{T\|P}(t \mid p)$	0.96	0.90	0.77	0.57	0.43
(n=33)	$\bar{S}^F_{T\|P}(t \mid p)$	0.91	0.82	0.67	0.43	0.28

fitted distributions can properly be considered as predictive. In order to assess the overall accuracy of the two approaches without the problem of having to define a suitable loss function (Henderson, 1995), the fitted survival probabilities were evaluated at the observed outcomes, ie we obtained $S_i = \hat{S}_{T|P}(t_i \mid p_i)$ and $S_i^F = \hat{S}^F_{T|P}(t_i \mid p_i)$. These will have standard uniform distributions if the fitted distributions are appropriate. The results are summarised in Table 6, which shows the proportion of observed values in intervals of length 0.25. Censored observations, which yield only upper bounds to S_i and S_i^F, are allocated under the assumption of a uniform distribution between the upper bound and zero. The frailty-conditional values are evenly distributed, those for the Weibull-conditional a little less so, with rather too few values in excess of 0.75, implying that the Weibull conditionals on average tend to underestimate survival a little over the first quartile. Kolmogorov-Smirnov tests however indicate no significant departures from uniformity for either S_i or S_i^F.

The frailty and Weibull conditionals were very similar for many individual patients but for others they were sometimes very different. Figure 6 gives six examples. The top two are selected randomly from patients with predicted value 1 month, the next two chosen randomly from those predicted to live 6 months and the bottom row has two from those with prediction 24 months. Some differences were possible at all values of P but the most substantial differences occurred when P was large. Here the two approaches lead to very different conditional distributions, with the Weibull-conditional invariably giving lower survival over the observed range than the frailty-conditional, which is consistent with the results of Section 3. Table 7 for instance shows mean values of $\hat{S}_{T|P}(t \mid p)$ and $\hat{S}^F_{T|P}(t \mid p)$ for patients with very low, intermediate or very high predictions. For low and medium P the differences are of no practical consequence but when P is large the Weibull-conditional gives substantially lower estimated survival probabilities than the frailty-conditional.

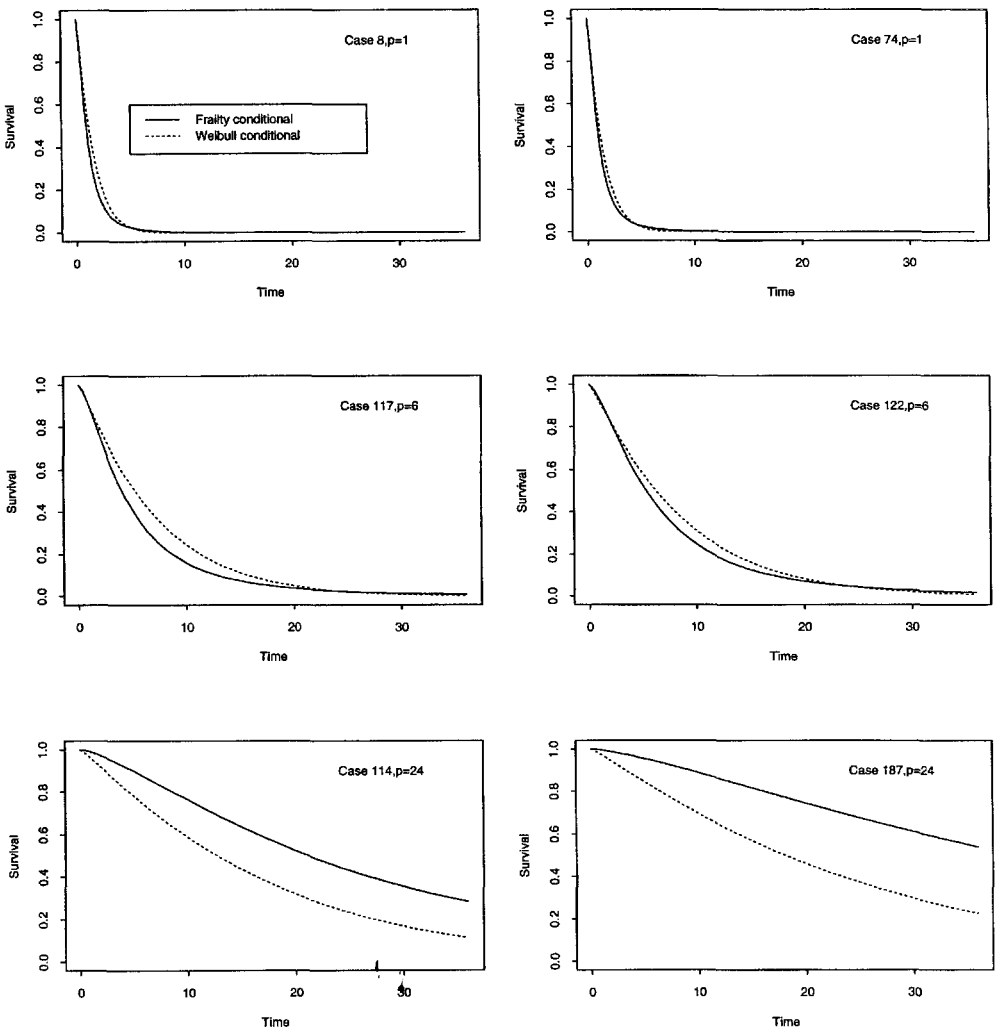


Figure 6. Examples of conditional fits.

5. Discussion

When interest is in the conditional distribution of one survival time T given the value of another, P, it would be convenient if P could simply be included as a covariate in a univariate model for T. This analysis has explored the consequences of such an approach when the true bivariate distribution follows the commonly assumed gamma frailty model and has indicated that the misspecified model leads to underestimation of early survival probabilities, and overestimation later. Sometimes the differences are relatively minor but at other times, particularly when P is large, practically important differences occur. The fitted Weibull conditional in these circumstances will often have lower mean and variance than a true frailty conditional, especially when the data include censoring. Thus the direct method is unsafe for data which can be assumed to follow the gamma frailty model, which continues to be many authors' first-line approach when analysing bivariate survival data. Extension of the results to allow more general distributions for the frailty term, including the other commonly made assumption of positive stable Z (Hougaard, 1987), is currently being investigated.

Another reason for avoiding the direct approach to modelling a conditional distribution concerns the interpretation of covariate effects. As shown in Section 2, the fitted coefficients will lie, asymptotically, between $\underline{\tilde{\beta}_t}$ and $0.5(\underline{\tilde{\beta}_t} - \underline{\tilde{\beta}_p})$ depending upon the strength of association ($\lambda = \infty$ to 0). In many applications the marginals for T and P will be similar, so that $\underline{\tilde{\beta}_t} \approx \underline{\tilde{\beta}_p}$, and consequently most covariates will apparently have little effect on conditional survival, whereas in fact their effect is masked by that of P. As an aside, we note that a further disadvantage of using P as a covariate in a univariate model for T is the difficulty of including censored values for P, whereas this is straightforward under a bivariate model.

Finally, as pointed out by one of the referees, a semiparametric Cox model is perhaps more likely to be used in practice than a Weibull model for the misspecified conditional distribution. Thus it would be useful to extend the investigation to incorporate this possibility. If the bivariate model of Section 2.1 applies but a Cox model with $\log(P)$ as covariate is fitted by maximum partial likelihood, then it can be shown that in the absence of other covariates the estimated coefficient of $\log(P)$ is asymptotically consistent when there is no censoring to the solution α_0 of

$$C + \psi(\lambda) + \psi(\tfrac{\alpha_0}{\delta} + 1) - \psi(\lambda - \tfrac{\alpha_0}{\delta}) + \frac{1}{\lambda} = 0$$

Specimen values, marked "Cox", are included in Figure 2. The solution is close to, but not identical with, the corresponding solution for the Weibull fit. The effect of incorporating censoring or other covariates is not yet known.

Acknowledgements

The author is grateful to Margaret Jones and the Thoracic Group of the Yorkshire Regional Cancer Organisation for permission to use the data of Section 4, and for the helpful comments of Jonathan Tawn and the referees.

References

1. O.O. Aalen," Heterogeneity in Survival Analysis," *Statistics in Medicine* vol 7, pp. 1121–1137, 1988.
2. O.O. Aalen,"Effects of Frailty in Survival Analysis," *Statistical Methods in Medical Research* vol 3, pp. 227–243, 1994.
3. O.O. Aalen, E Bjertness and T.Sønju,"Analysis of Dependent Survival Data Applied to Lifetimes of Amalgam Fillings," *Statistics in Medicine* vol 14, pp. 1819-1829, 1995.
4. J.E.Anderson and T.A.Louis, "Survival Analysis Using a Scale Change Random Effects Model," *Journal of the American Statistical Association* vol 90, pp. 669–679, 1995.
5. D.Clayton,"Some Approaches to the Analysis of Recurrent Event Data," *Statistical Methods in Medical Research* vol 3, pp. 244–262, 1994.
6. D.Clayton and J.Cuzick,"Multivariate Generalizations of the Proportional Hazards Model (with Discussion)," *Journal of the Royal Statistical Society Series A* vol 148, pp. 82–117, 1985.
7. I.S.Gradshteyn and I.M.Ryzhik, *Tables of Integrals, Series and Products, 5th ed.*, Academic Press: London, 1994.
8. R.Henderson, "Problems and Prediction in Survival Analysis," *Statistics in Medicine* vol 14, pp. 161–184, 1995.
9. R.Henderson and M.Jones, "Prediction in Survival Analysis: Model or Medic?" *Lifetime Data: Models in Reliability and Survival Analysis, ed N.P.Jewell, A.C.Kimber, M-L.T.Lee and G.A.Whitmore.* Kluwer: Dordrecht, 1995.
10. P.Hougaard, "Modelling Multivariate Survival," *Scandinavian Journal of Statistics* vol 14, pp. 291–304, 1987.
11. P.Hougaard, "Frailty Models for Survival Data," *Lifetime Data Analysis* vol 1, pp. 255–273, 1995.
12. P.Hougaard, B.Harvald and N.V.Holm, "Measuring the Similarities Between the Lifetimes of Adult Danish Twins Born Between 1881–1930," *Journal of the American Statistical Association* vol 87, pp. 17–24, 1992.
13. P.Hougaard, P Myglegaard and K.Borch-Johnsen, "Heterogeneity Models of Disease Susceptibility, with Application to Diabetic Nephropathy," *Biometrics* vol 50, pp. 1178–1188, 1994.
14. C.A.McGilchrist,"REML Estimation for Survival Models with Frailty," *Biometrics* vol 49, pp. 221–225, 1993.
15. A.Pickles and R.Crouchley, "Generalizations and applications of frailty models for survival and event data," *Statistical Methods in Medical Research* vol 3, pp. 263–278, 1994.
16. J.H.Shih and T.A.Louis, " Assessing Gamma Frailty Models for Clustered Failure Time Data," *Lifetime Data Analysis* vol 1, pp. 205–220, 1995.
17. J.T.Wassell, G.W.Kulczycki and E.S.Moyer, "Frailty Models of Manufacturing Effects," *Lifetime Data Analysis* vol 1, pp. 161–170, 1995.