Variance Components Models for Survival Data

Jørgen Holm Petersen[12] Per Kragh Andersen[12] Richard D. Gill[3]

[1] Department of Biostatistics, University of Copenhagen,
Blegdamsvej 3, DK 2200 Copenhagen N, Denmark

[2] Danish Epidemiology Science Centre
[3] Department of Mathematics, University of Utrecht,
PO Box 80010, NL 3508 TA Utrecht, The Netherlands

Abstract

Extensions of the Cox proportional hazards model for survival data are studied where allowance is made for unobserved heterogeneity and for correlation between the life times of several individuals. The extended models are frailty models inspired by Yashin et al. (1995). Estimation is carried out using the EM algorithm. Inference is discussed and potential applications are outlined, in particular to statistical research in human genetics using twin data or adoption data, aimed at separating the effects of genetic and environmental factors on mortality.

Keywords: Censored survival data, heterogeneity, correlated frailty, correlated life times, semiparametric models, EM algorithm.

1 Introduction.

This paper is motivated by recent and ongoing studies in human genetics whose aim is to separate the effects of genetic and environmental factors on longevity or on (age-dependent, cause-specific) mortality. Data consists of life-times and causes of death of members of 'family units' in a large sample of the latter. The family units can be of quite elaborate structure. Adoption and twin registers are favourite sources of data for such studies.

Two statistical traditions meet here. Variance components models based on normal distributions are widely used in genetics, especially in animal breeding, to study quantitative traits of individuals. In the present context typically the logarithm is taken of the survival time followed by a traditional random effects linear model analysis. Censoring of survival times is hard to take account of in such an analysis, and it is difficult to motivate that genetic and environment effets act multiplicatively on total lifetime.

The other statistical tradition is modern surival analysis, with its emphasis on dynamic modelling of the hazard rate of independent individuals. Censoring and delayed entry (left truncation) does not complicate the analysis. However the analysis of *dependent* survival times, especially when subject to complicated patterns of dependence, is so far not much developed.

Here we propose a synthesis of these traditions by superimposing an additive variance components type structure on so-called multiplicative gamma frailty models for survival analysis. These frailty models were originally developed to allow for unobservable heterogeneity (missing covariates) though their use has remained somewhat controversial due to identifiability issues. Later they have been extended to model simple patterns of dependence, but in perhaps a rather (too) simplistic way.

The next step, structuring the dependence structure, was first and only recently taken by Yashin et al. (1995); see also Iachine (1995). At first sight it seems curious to combine frailty components additively, which then act multiplicatively on individual hazard rates. Initially the only advantage is consistency with earlier proposed models for independent individuals (since adding independent gammas with the same shape results again in a gamma distribution). Our new contribution is to show that the model also has statistical advantages: despite its elaborate nature, it is amenable to more or less straight-forward statistical analysis (non-parametric maximum likelihood estimation) smoothly extending the highly successful established techniques of survival analysis (in particular, the Cox regression model). We show how the EM algorithm can be applied to this situation, involving recursive calculation formulas for expected components of frailty in the E-step.

The end result, extending in a smooth way ordinary survival analysis methodology and incorporating the very intuitively appealing variance components models, seems like a clear and better alternative to the traditional analysis of life-time data in genetics. Although not based on stringent genetic theory, we believe that this approach represents a step forward in the direction of learning something about the data and underlying genetic and environmental effects.

The paper is structured according to the historical development in survival analysis outlined above. In classical regression analyses of survival data (e.g., Cox, 1972; Andersen,

Borgan, Gill and Keiding, 1993, Chapter VII) the assumptions usually include that: (a) the model is correct in the sense that all relevant covariates are registered and included in the model, and (b) all individuals under study have life times which are statistically independent. In Section 2 an introduction to classical survival analysis is given with special emphasis on the so-called counting process approach. Section 3 discusses *frailty models* for heterogeneous populations, i.e., models where (a) above is relaxed while in Section 4 we study frailty models for correlated survival data (dealing with (b)). Some final remarks on further developments and conclusions are given in Section 5 and 6.

2 Introduction to survival analysis.

Survival analysis deals with life times or, more generally, with periods elapsed from some initial event at time 0 (like birth, start of treatment or employment in a given job) to some terminal event of interest (like death, relapse or disability pension). Thus the basic data would ideally be independent non-negative random variables $(T_i, i = 1, \ldots, n)$. What distinguishes survival times from other kinds of data, however, is the inevitable presence of incomplete observations. Often, practical restrictions prevent the observation of the terminal event of interest in every individual, i, in the sample in which case the available piece of information is a *right-censoring* time U_i, i.e., a period elapsed in which the event of interest has not occurred (e.g., a patient has survived until U_i). Thus, a generic survival data sample includes $((\widetilde{T_i}, D_i), i = 1, \ldots, n)$ where $\widetilde{T_i}$ is the smaller of T_i and U_i and D_i is the indicator, $I(T_i \leq U_i)$, of not being censored. Another kind of incompleteness frequently encountered in practice is *left-truncation* where individual i is only included in the sample conditionally on having "survived" without the terminal event till some given entry time $V_i \geq 0$, i.e., individual i is only observed from V_i and onwards.

The distribution of T_i may be described by the *survival function*

$$S_i(t) = \Pr(T_i > t) = \exp(-A_i(t)),$$

where

$$A_i(t) = \int_0^t \alpha_i(u)du$$

is the integral over $[0, t\ )$ of the *hazard function*

$$\alpha_i(t) = \lim_{\Delta t \to 0} \Pr(T_i \leq t + \Delta t \mid T_i > t)/\Delta t.$$

Due to the dynamical nature of survival data, the latter characterisation of the distribution via the hazard function is often more convenient. (Note that $\alpha_i(t)\Delta t$ when $\Delta t > 0$ is *small* is approximately the conditional probability of i "dying" just after time t given "survival" till time t.) Also, $\alpha_i(t)$ is the basic quantity in the *stochastic process approach* to survival analysis, see, e.g., Andersen et al. (1993) which we shall also adopt in this paper.

In this approach the survival data are represented as *counting processes* $(N_i(t), i = 1, \ldots, n)$ where

$$N_i(t) = I(\widetilde{T_i} \leq t, D_i = 1)$$

counts 1 at $\widetilde{X}_i$ if i is not censored and otherwise $N_i(t) = 0$ throughout. The dynamics of $N_i(t)$ is described by its (random) *intensity process* which is a product

$$\lambda_i(t) = \alpha_i(t)Y_i(t)$$

of the hazard function and the random process

$$Y_i(t) = I(\widetilde{T}_i \geq t)$$

(or $Y_i(t) = I(\widetilde{T}_i \geq t > V_i)$ if there is left-truncation) indicating whether i is observed to be *at risk* just before time t. The interpretation of the intensity process is that

$$\lambda_i(t)\Delta t \approx E(dN_i(t) \mid \mathcal{F}_{t-}),$$

the conditional expectation of the jump size $dN_i(t)$ at time t given the observed "history" $\mathcal{F}_{t-}$ of individual i in $[0, t\ )$ (and possibly others if there is dependence among the individuals in the sample). Mathematically, $(\mathcal{F}_t)_{t\geq 0}$ is a *filtration*, i.e., an increasing right-continuous family of σ-algebras and both $\boldsymbol{N}(\cdot) = (N_1(\cdot), \ldots, N_n(\cdot))$ and $\boldsymbol{Y}(\cdot) = (Y_1(\cdot), \ldots, Y_n(\cdot))$ are adapted to $(\mathcal{F}_t)$.

Regression models may be constructed by including *covariates* $(\boldsymbol{X}_i, i = 1, \ldots, n)$ (which may be *time-dependent*) in $(\mathcal{F}_t)$ and letting the intensity process depend on these:

$$\lambda_i(t) = \alpha_i(t; \boldsymbol{X}_i)Y_i(t).$$

Frequently, the *Cox proportional hazards* model is studied where the hazard function

$$\alpha_i(t; \boldsymbol{X}_i) = \alpha_0(t) \exp(\beta^T \boldsymbol{X}_i)$$

is a product of an unknown *baseline hazard* $\alpha_0(t)$ common to all individuals and a relative risk $\exp(\beta^T \boldsymbol{X}_i)$ where the covariates $\boldsymbol{X}_i$ enter via a vector of unknown *regression coefficients*.

3 Extending the Cox-model: The Frailty Model

The Cox proportional hazards model has enjoyed a widespread acceptance in statistical applications. This is due in part to the intuitively appealing interpretation of the hazard function, but also because estimation and inference is mathematically feasible.

In biostatistics, biological variation between subjects, whether human or animal, can be considerable. This variation is often partially accounted for by means of extensive covariate histories on the subject level, but even after having included this information in the analysis, a major source of variation often remains unaccounted for.

In models based on an assumption of Gaussian responses, such unobservable variation may be included by means of variance components. With only a single measurement per subject, it is usually considered part of the measurement error, i.e., as a part of the random

variation of the experiment. In this case, the extra variation is not a serious concern. The usual estimators will be consistent and unbiased if the unobserved variation is independent of the observed covariates. The only price is larger variance estimates and thus wider confidence intervals which will tend to favor the null hypothesis.

In survival analysis, however, this unobserved variation if ignored, can lead to serious bias in both parameter estimates and in the estimate of the hazard rate, see Bretagnolle and Huber-Carol (1988). The reason for the bias is the very property that makes the survival analysis models so appealing: the time-dependent hazard rate. If some individuals are at higher risk to experience an event due to some unobserved variables, then the individuals remaining at risk tend to be a selected group with an associated lower risk. An estimate of the hazard rate without taking into account the unobserved variables will therefore underestimate the true hazard to a greater and greater extent as time increases.

This stresses the point that caution should be exercised when interpreting the hazard rate. It is a result of two different sources of variation, one within subjects reflecting the risk changing over time of a given subject, and the other, the selection of individuals prone to failure, reflecting the variation among subjects. If both of these sources of variation are present and we do not include them in our analysis, both the interpretability of the hazard rate as the evolution of individual risk over time, as well as the estimates of say, treatment effects, are at best obscured and at worst seriously biased. These aspects of the effects of selection are discussed in more detail by, e.g., Vaupel and Yashin (1985) and Aalen (1994).

3.1 Frailty Models

To address the issue of variation due to unobserved variables Vaupel et al. (1979) introduced a random effect into a survival model framework. They introduced the term frailty model and applied the model in a demographic setting to adjust/account for population heterogeneity.

The frailty model assumes a proportional hazard model conditioned on the random effect. Specifically, the hazard rate $\alpha_i(t)$ of an individual depends on an unobserved random variable Z_i, acting multiplicatively on a baseline hazard $\alpha(t)$, i.e.,

$$\alpha_i(t; Z_i) = Z_i \alpha(t). \tag{1}$$

Here Z_i is considered as a random variable varying over the population of individuals. It can be thought of as a *frailty* or accident-proneness which increases the susceptibility to failure.

The model is rather simple in that all individuals, apart from a constant Z_i, are assumed to follow the same mortality pattern expressed by the baseline hazard. However, bearing in mind the success of the semi-parametric models in survival analysis, notably the Cox model, and the above mentioned possibilities of bias due to unobserved covariates, this type of model may yield considerable insight as to the effect on estimation of parameters of interest due to unobserved variables.

Estimation and inference in this model requires some structure, either on the baseline

hazard or on the distribution of the unobserved frailty. Choosing the latter yields a natural extension of the Cox model with its non-parametric baseline hazard. In this setting, we note that the unobserved frailty enters the regression part of the hazard function in exactly the same manner as the observed covariates, i.e.,

$$\alpha(t; \boldsymbol{X}_i, Z_i) = \alpha(t) \exp(\boldsymbol{\beta}^T \boldsymbol{X}_i + \log Z_i). \tag{2}$$

It would seem reasonable to choose a log-normal distribution for the frailty since this would correspond to a normally distributed covariate. This however turns out less useful than a gamma distribution which has similar properties to the log-normal, i.e., it is unimodal and right skewed, and for which analytical expressions for the likelihood function are readily available. It is shown by Elbers and Ridder (1982), see also Kortram et al. (1995), that the model specified by (1) is identifiable with any finite mean frailty distribution. In the case of survival data, that is, when each individual can experience at most one event, we need also covariates entering the hazard function in a specified way, e.g. (2). Without covariates and with maximally one event per individual the model is not identified.

Using the innovation theorem, Brémaud (1981), the population hazard $\mu(t)$ at a given instant from (1), is found to be $\mu(t) = E\{Z \mid T > t\}\alpha(t)$. This demonstrates that the population hazard is the average hazard of the surviving individuals. A further requirement of the frailty distribution which would allow for left truncated survival times, is that this conditional distribution stays within the same family of distributions. This is the case for the gamma distribution as will be seen later.

The choice of the gamma distribution has been discussed and criticized, sometimes vigorously, by, e.g., Oakes (1989) and in a series of papers by Hougaard, see Hougaard (1995) and references therein. Alternatives based on the positive stable distributions are theoretically appealing but harder to analyse in practice, and are not discussed further here.

3.2 Likelihood and Estimation via the EM-algorithm

Our presentation is based on counting processes as introduced in Section 2, and follows to a large extent Nielsen et al. (1992). We describe the frailty model in its simplest semi-parametric form with a frailty variable and a non-parametric baseline hazard. The identifiabilty problem of the model, in the case of survival data with at most one event per individual, would as mentioned in the previous section require covariates. Such extentions of the model to include explicitly e.g Cox-type covariates or stratified baseline hazards, is fairly straightforward and is only discussed briefly in the following.

Let $\boldsymbol{N} = (N_i; i = 1, \ldots, n)$ be a multivariate counting process with intensity process $\boldsymbol{\lambda} = (\lambda_i; i = 1, \ldots, n)$ satisfying

$$\lambda_i(t) = Z_i Y_i(t) \alpha(t) \tag{3}$$

where Y_i is an observed predictable process, α is an unknown baseline hazard function, and the Z_i's are unobserved random variables, independently drawn from a gamma(ν, η)

distribution. In the following, we let $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$.

A restriction which is needed to make $\alpha(t)$ identifiable may be imposed by arbitrarily setting the mean frailty of individuals to one, i.e., by setting $\nu = \eta$. This allows an interpretation of $\alpha(t)$ as the hazard rate of an "average" subject. The remaining parameter η determines the variance, and the squared coefficient of variation, of Z_i: $V\{Z_i\} = \eta^{-1}$.

As discussed in detail by Nielsen et al. (1992), we need apart from the usual assumption of independent censoring also that the censoring is noninformative of $\boldsymbol{Z}$. With these assumptions and if $\boldsymbol{Z}$ were observed, valid inference could be drawn from the complete data (partial) likelihood

$$\prod_i \Big\{ p(z_i;\nu,\eta) \prod_t (z_i Y_i(t) dA(t))^{\Delta N_i(t)} \exp\Big(-z_i \int_0^\tau Y_i(s) dA(s)\Big) \Big\} \tag{4}$$

where p is the gamma density $p(z;\nu,\eta) = \eta^\nu z^{\nu-1} e^{-\eta z}/\Gamma(\nu)$.. Here, τ denotes the end of the observation period and A is the integrated baseline hazard $\int \alpha$. Now direct integration of (4) with respect to the unobserved frailty variable, yields the observed data (partial) likelihood

$$L(\alpha,\eta) = \prod_i \Big\{ \frac{\eta^\nu}{\Gamma(\nu)} \frac{\Gamma(\nu + N_i(\tau))}{(\eta + \int_0^\tau Y_i(s) dA(s))^{\nu + N_i(\tau)}} \prod_t (Y_i(t) dA(t))^{\Delta N_i(t)} \Big\}. \tag{5}$$

This likelihood is a function of the unknown entities, i.e., the frailty variance $(1/\eta)$ and the (cumulative) baseline hazard. If a parametric baseline hazard is specified, this is usually a function of a relatively small number of parameters and numerical maximization frequently fairly uncomplicated. In the semi-parametric case, that is, with a nonparametric baseline hazard, this approach is less useful.

It turns out useful, however, to consider the statistical problem of maximum likelihood estimation of $A = \int \alpha$ and η as an incomplete or missing data problem. The complete but unobserved data consist of $\boldsymbol{Z}$, $\boldsymbol{N}$ and $\boldsymbol{Y}$. The incomplete and actually observed data in turn, consist of the last two components.

In the complete data problem, that is, had the frailties been observed, we would just absorb the frailty variables into the random part $Y_i(t)$ of the intensity process, and the model reduces to a standard counting process model. The cumulative baseline hazard could in this case be estimated by the usual Nelson-Aalen estimator. In the regression setting, i.e., with Cox-type covariates, regression parameters would be estimated by the Cox partial likelihood. This is described in further detail in the following section.

That the complete data problem at hand is so simple as opposed to the incomplete data problem, makes the EM algorithm an interesting alternative to the high dimensional numerical maximization, Gill (1985).

The EM-algorithm, Dempster, Laird and Rubin (1977), is a general algorithm for maximum likelihood estimation in incomplete data problems. The algorithm consists of two steps—an Expectation step and a Maximization step. In the E-step, we calculate for a fixed set of parameters $A^{(r)}$ and $\eta^{(r)}$, the conditional expectation of the complete data log-likelihood given the observed data $\boldsymbol{N}$ and $\boldsymbol{Y}$. This expectation is denoted

$Q(A, \eta \mid A^{(r)}, \eta^{(r)})$ where A and η are the parameters in the log-likelihood, and $A^{(r)}$ and $\eta^{(r)}$ are the parameters in the conditional distribution. Now, Q is a function of A and η which may be maximized. This maximization is the M-step and results in values $A^{(r+1)}$ and $\eta^{(r+1)}$. It may be shown that this scheme increases the observed data likelihood and iterating until convergence thus yields a (local) maximum of the observed likelihood function which was also the original goal. Informally, the M-step is a ordinary maximization in the complete data problem. The E-step in turn, consists of calculating the proper statistics for carrying out the M-step.

Taking logarithms in (4) and rearranging yields for each individual

$$\sum_t \Delta N_i(t) \log(dA(t)) - z_i \int_0^\tau Y_i(s) dA(s) + \sum_t \Delta N_i(t) \log(z_i) + \log(p(z_i; \nu, \eta)). \qquad (6)$$

The first two terms contain the unknown hazard, but not the frailty parameters, whereas the converse is the case for the last two terms. This, in a sense orthogonal parameter space structure, is exploited in the estimation. Consider the case where η is fixed throughout the entire EM iteration scheme. Then, since z_i enters only linearly in (6) when viewed as a function of A, the E-step reduces to the calculation of the conditional expectation of the frailties given the observed data $\boldsymbol{N}$ and $\boldsymbol{Y}$. This is simple since the complete data likelihood (4) as a function of $\boldsymbol{z}$ is proportional to

$$\prod_i \left\{ z_i^{\nu + N_i(\tau) - 1} \exp(-z_i \eta - z_i \int_0^\tau Y_i(s) dA(s)) \right\}.$$

Thus, conditional on the data, the Z_i are still independent and gamma distributed but now with parameters $\nu + N_i(\tau)$ and $\eta + \int_0^\tau Y_i(s) dA(s)$. The conditional expectation $\widehat{Z}_i$, then equals

$$\widehat{Z}_i = \frac{\nu + N_i(\tau)}{\eta + \int_0^\tau Y_i(s) dA(s)}. \qquad (7)$$

The M-step now amounts to calculating the Nelson-Aalen estimate of the cumulative hazard as if $\boldsymbol{Z}$ had been observed and was equal to $\widehat{\boldsymbol{Z}}$, i.e.,

$$\widehat{A}(t) = \int_0^t \frac{dN_\cdot(s)}{\sum_i \widehat{Z}_i Y_i(s)}. \qquad (8)$$

To estimate the frailty variance η which was fixed throughout the EM-iteration, we maximize numerically the observed data likelihood, now for fixed cumulative hazard. This is a one-dimensional optimization problem and usually fairly straightforward. Now, for this new estimate of the frailty variance which also increased the observed data likelihood, we go through the EM-steps and iterate until convergence. After convergence, and since the observed data likelihood in each step is increased, we find a stationary point of the observed data likelihood, and under further conditions, it actually maximizes the likelihood.

This estimation scheme based on the EM-algorithm is slightly different from that described by Nielsen et al. (1992). They estimate regression parameters and baseline hazards

using the EM-algorithm for given value η at a series of values for η. They then choose as a global maximum likelihood estimate the η with the highest value of the observed data likelihood. This works well when the parameter space is one-dimensional, that is, when the only unknown frailty parameter is the frailty variance. However, it is not a very elegant or efficient procedure when the parameter space increases in dimension. This is relevant in connection with the litter frailty model presented in the next section.

Similar EM-type schemes have been suggested, Clayton and Cuzick (1985), and Self and Prentice (1986). A conceptually more direct approach discussed by Klein (1992), is to use the EM-algorithm also to estimate the frailty variance. This, however, involves the digamma function when calculating the conditional expectation of $\log(Z_i)$ given the observed data $\boldsymbol{N}$ and $\boldsymbol{Y}$. A Bayesian approach to frailty models using Gibbs sampling may be found in Clayton (1991).

Inference in frailty models is not yet completely resolved although Murphy (1994, 1995) has shown the existence, consistency and asymptotical normality of the estimators. In applications of frailty models, it has been assumed that the usual asymptotic results concerning likelihood ratio tests and confidence intervals, are valid. This is supported by simulation results of Nielsen et al. (1992). Variance estimates can be obtained from the observed data log-likelihood, Andersen et al. (1995) or, staying within the general EM-algorithm framework, using an approach suggested by Louis (1982).

4 A Multivariate Extension of the Cox model

So far, we have focused on the frailty models as a way of dealing with possible heterogeneity or overdispersion due to unobserved covariates. Another aspect of these models, and of random effects models in general, is to use them to model statistical dependence, e.g., Clayton (1978). This dependence could be between recurrent or different events for the same individual, e.g., onset of disease and subsequent death, or it could be between the same event for different individuals, e.g. onset of disease in twins, litters or families.

Traditionally, this type of multivariate counting process data has been dealt with by introducing the occurence times directly into the intensity processes. In this autoregressive approach, the correlated event times are thus modelled as if it were the actual events that change the intensity for subsequent failures. Although this may be the case in some situations, e.g. in a competition for a limited amount of food, it is at other times not a relevant modelling approach. In studies of, for example, the life times of different families, it is not the life times themselves that have an effect on the death intensities of the remaining family members. Rather, the observed life times are correlated because of the genetic and/or environmental circumstances that make members of families more similar than randomly selected subjects from the general population. The regression approach does not satisfactorily account for this type of dependence whereas the random effects approach does.

The dependent or correlated event times may be modeled using a frailty model by letting several components of a multivariate counting process share the same frailty variable. Note

that we have already used this notion of correlated event times by letting a counting process with possibly several jumps share a frailty. By letting a multivariate counting process share the same frailty variable, we induce positive statistical dependence between the individual counting processes. The parameter (η) of the frailty distribution now acts as an association parameter since the event times become more strongly associated as the frailty variance η^{-1} increases. This type of model, although conceptually different, has the same likelihood expressions and estimation and inference is carried out as described in the previous section.

Modeling dependence in this way, however, sometimes restricts the joint distribution of the life times beyond what the data allow. If there is both association and heterogeneity present in the data, then the parameter (η) measures not only the association but also heterogeneity. Unless one is very confident with the model that is specified, this confounding of effects makes the frailty model as described here less suitable for modelling correlated event time data.

The effect of the confounding is illustrated by Vaupel et al. (1992). They find in two separate analyses of Danish monozygotic (MZ) and dizygotic (DZ) twins, a higher η and a steeper baseline hazard for monozygotic than dizygotic twins.

To understand this, we note that the marginal survivor function is found from the conditional by integration with respect to the frailty distribution. One finds that

$$P\{T > t\} = \frac{1}{(1 + \frac{1}{\eta}\int_0^t Y(s)dA(s))^{\eta}}. \tag{9}$$

So, if we were to assume the same baseline hazard function for MZ and DZ twins and the same marginal survival function (9) which is very reasonable and suported by empirical evidence, then η for MZ and DZ twins must also be identical. But there are literally hundreds of studies indicating that MZ twins with respect to many different endpoints are more closely associated than DZ twins. This is a drawback of the frailty model in the context of modeling correlated event times, and stems from the fact that individuals sharing the same frailty (and covariates) have exactly the same risk. As a way of circumventing this problem, Yashin et al. (1995) suggest decomposing the frailty of each twin in a pair into a sum of two independent frailties, one of which is shared by both twins. The following extended frailty model is based on this idea.

4.1 The Litter Frailty Model

Let $\boldsymbol{N} = ((N_{i1}(t), \ldots, N_{in_i}(t)); i = 1, \ldots, n)$ be a multivariate counting process. This set-up reflects the situation where n litters with n_i subjects in the i'th litter are observed. For simplicity, we assume the outcome to be time to death which means that each individual may experience only one event.

The aim is to specify intensity processes $\boldsymbol{\lambda}_i = (\lambda_{ij})$ where

$$\lambda_{ij}(t) = Z_i^{(j)} Y_{ij}(t) \alpha_h^{\theta}(t; \boldsymbol{X}_{ij}) \; ; \; i = 1, \ldots, n; j = 1, \ldots, n_i; h = 1, \ldots, k. \tag{10}$$

Here, $Y_{ij}(t)$ is predictable, and is usually just indicating whether the subject indexed by ij is at risk for an event at time $t-$, see Section 2. The function $\alpha_h^\theta(t; \boldsymbol{X}_{ij})$, which specfies the regression part of the model, is given by

$$\alpha_h^\theta(t; \boldsymbol{X}_{ij}) = \alpha_h(t) \exp(\boldsymbol{\beta}^T \boldsymbol{X}_{ij}) \tag{11}$$

where $\boldsymbol{X}_{ij}$ is a vector of covariates specific to subject ij and β is a p-vector of regression parameters. In this semi-parametric setting θ just equals β, otherwise it would also include the parameters of the baseline hazard. Although other choices of relative risk regression functions are possible, the exponential is convenient and makes explicit the connection to the traditional Cox-model.

The baseline hazard $\alpha_h(t)$ is indexed by h allowing for different baseline hazards for subjects in different strata given by e.g. sex. Note that ij uniquely determines each subject, and thus also to which stratum that subject belongs. Omitting the indices in $h = h(i,j)$ should not cause confusion. The baseline hazards are unknown and to be estimated. Furthermore, $Z_i^{(j)}$ is a random component which is specific to each subject in each litter. The random variables within litters are correlated through the following construction

$$\begin{array}{rcl} Z_i^{(1)} & = & Z_{i0} + Z_{i1} \\ \vdots & & \vdots \\ Z_i^{(n_i)} & = & Z_{i0} + Z_{in_i} \end{array} \tag{12}$$

where $Z_{i0}, Z_{i1}, \ldots, Z_{in_i}$ are independent gamma distributed random variables with parameters $(\nu, \eta), (\nu^*, \eta), \ldots, (\nu^*, \eta)$, respectively. The common scale parameter (η^{-1}) implies that $Z_i^{(j)}, j = 1, \ldots, n_i$ have gamma distributions which as with the usual 1-dimensional frailty model will turn out to be convenient. As previously, a restriction is needed to make $\alpha_h(t)$ identifiable and this is done by restricting the mean frailty of individuals to one, i.e., by setting $\eta = \nu + \nu^*$. This allows an interpretation of $\alpha_h(t)$ as the underlying intensity, in the given strata, of an "average" subject with $\boldsymbol{X}_{ij} = \boldsymbol{0}$. Note also how the common frailty Z_{i0} will induce correlation between the intensities of subjects within litters and thus also a correlation between their life times. This correlation could be attributable to their shared genes as well as to the common environment to which they are exposed. The common shape parameter of $Z_{i1}, \ldots, Z_{in_i}$ models the possible heterogeneity between individuals even after having accounted for the common genes and the common environment. This could reflect the non-shared environment to which the individuals are exposed. Details of this model may be found in Petersen (1995).

It is worthwhile to look into the possible uses and special cases of this model. Some of these models reduce to well-known one-dimensional frailty models.

- one-sample problem

$$\lambda_{ij}(t) = Z_{i0} Y_{ij}(t) \alpha(t) \; ; \; i = 1, \ldots, n; j = 1, \ldots, n_i. \tag{13}$$

In model (13) subjects are correlated through Z_{i0}. This could be relevant as a model for estimating $\alpha(t)$, i.e., the distribution of life times, in a litter-matched study. The life times of subjects within litters are identically distributed, but not independent.

- k-sample problem

$$\lambda_{ij}(t) = Z_{i0}Y_{ij}(t)\alpha_h(t)\ ;\ i = 1,\ldots,n; j = 1,\ldots,n_i; h = 1,\ldots,k. \tag{14}$$

Model (14) is an extension of (13) in that it allows for different baseline hazards in strata defined by h. This type of model could be considered for family study data, e.g. with $k = 4$ corresponding to father, mother, male and female offspring. Apart from the statistical advantages of this model as opposed to estimating the life time distributions based on subjects in each stratum, this approach would, if the model is true, allow for an assessment of inter-family variability.

- k-sample problem with proportional hazards

$$\lambda_{ij}(t) = Z_{i0}Y_{ij}(t)\exp(\boldsymbol{\beta}_h)\alpha(t)\ ;\ i = 1,\ldots,n; j = 1,\ldots,n_i; h = 1,\ldots,k. \tag{15}$$

Model (15) is another extension of model (13) and relevant when each of the subjects in the litter undergoes one of k treatments, and when treatments act multiplicatively on the baseline hazard. This model is readily extended to a more general regression model, i.e., to include covariates such as age, smoking habits etc.

- overdispersion model

$$\lambda_{ij}(t) = Z_{ij}Y_{ij}(t)\alpha(t)\ ;\ i = 1,\ldots,n; j = 1,\ldots,n_i. \tag{16}$$

Model (16) is a formulation of the original frailty model, used by Vaupel et al. (1979), to model heterogeneity due to unobserved covariates. Each subject has its own associated frailty component acting multiplicatively on the baseline hazard. Frailties (and life times) associated with different subjects are independent. Note that the notion of litters in this setup has no meaning. The indices ij are kept only to conform with the previous notation and the model is simply the one introduced in Section 3.1.

- twin model

$$\begin{aligned} \lambda_{i1}(t) &= (Z_{i0} + Z_{i1})Y_{i1}(t)\alpha_h(t) \\ & \qquad\qquad\qquad\qquad i = 1,\ldots,n\ ; \\ \lambda_{i2}(t) &= (Z_{i0} + Z_{i2})Y_{i2}(t)\alpha_h(t)\ . \end{aligned} \tag{17}$$

Model (17), suggested by Yashin et al. (1995), has a fixed litter size of 2. The model applies to classical twin studies and the frailties here have the interpretation of representing genetic and shared environmental effects (Z_{i0}) and non-shared environmental effects (Z_{i1}, Z_{i2}). The model allows for different baseline hazards for male and female twins.

- extended litter model

$$\begin{aligned} \lambda_{i1}(t) &= (Z_{i0} + Z_{i1})Y_{i1}(t)\alpha(t)\ ; \\ \vdots & \qquad\qquad \vdots \qquad\qquad i = 1,\ldots,n; j = 1,\ldots,n_i\ ; \\ \lambda_{in_i}(t) &= (Z_{i0} + Z_{in_i})Y_{in_i}(t)\alpha(t)\ . \end{aligned} \tag{18}$$

This model is an extension of the model (13) for the one-sample problem along the lines of the twin model. As opposed to model (13) it allows for heterogeneity among subjects within litters. Appart from being a seemingly more reasonable model in a litter study context, it also has more operational value, since it may be used to test the goodness of fit of a standard one-dimensional frailty model.

4.2 Likelihood Construction

The estimation of η and α is carried out by maximization of the likelihood function based on the observed data $(\boldsymbol{N}, \boldsymbol{Y}) \mid \boldsymbol{L}$ where $\boldsymbol{L}$ denotes the litter sizes. The parameters associated with these litter sizes are usually not of interest.

As in the one-dimensional frailty model, we need to assume that conditional on $(\boldsymbol{Z}, \boldsymbol{L})$, censoring is independent, and that conditional on $\boldsymbol{Z}$, censoring is non-informative of $\boldsymbol{Z}$. A similar assumption must be made concerning the litter size distribution, i.e., that the litter sizes must not depend on the frailties.

With these assumptions, we can carry out steps similar to the one-dimensional frailty model. The complete data (partial) likelihood is a product of the likelihood contributions from each litter, i.e.,

$$\prod_i \left\{ \prod_{j=1}^{n_i} \left(p(z_{ij}; \nu^*) \prod_t \lambda_{ij}(t)^{\Delta N_{ij}(t)} \right) \exp\left(- \sum_{j=1}^{n_i} z_i^{(j)} \Lambda_{ij} \right) p(z_{i0}; \nu) \right\} \tag{19}$$

where $\Lambda_{ij} = \int_0^\tau Y_{ij}(s) \alpha_h^\theta(s; \boldsymbol{X}_{ij}) ds$ is the cumulative intensity process evaluated at τ.

Since individuals across litters are independent, we need only carry out calculations for one litter and therefore omit the litter index i in the following. Integrating out all the frailties that are particular to each individual $(z_{ij}; j = 1, \ldots, n_i)$ yields a polynomial in z_{i0} of an order depending on the number of deaths in the litter. For a litter with l subjects, the likelihood contribution from m deaths where $m = 0, \ldots, l$, is given by

$$\prod_{j=1}^{m} \alpha_h^\theta(t_j; \boldsymbol{X}_j) \prod_{j=1}^{l} \left(\frac{\eta}{\Lambda_j + \eta} \right)^{\nu^*} \sum_{j=0}^{m} C_j^{(m)} \frac{\Gamma(\nu + j)}{\Gamma(\nu)} \frac{\eta^\nu}{(\Lambda_\cdot + \eta)^{\nu + j}} \tag{20}$$

where $C_r^{(m)}$ is defined recursively by

$$\begin{aligned} C_r^{(r)} &= 1 \\ C_j^{(r)} &= C_{j-1}^{(r-1)} + c_r C_j^{(r-1)} \; ; \; j = 1, \ldots, r-1 \\ C_0^{(r)} &= c_r C_0^{(r-1)} \end{aligned} \tag{21}$$

starting with $C_0^{(0)} = 1$, and where $c_r = \frac{\nu}{*} / (\Lambda_r + \eta)$; $r = 1, \ldots, l$. Dot denotes summation over index j.

Now, likelihood estimation and inference could be carried out assuming an appropriate parametric form of the baseline hazard. However, we can go through exactly the same steps as for the one-dimensional frailty model assuming a non-parametric baseline hazard.

Taking logarithms in (19), replacing $\lambda_{ij}(t)$ by $z_i^{(j)} Y_{ij}(t) dA_h^\theta(t)$ and rearranging yields the following log-likelihood contribution for the i'th litter

$$\begin{array}{lcl} \sum_{j=1}^{n_i} N_{ij}(\tau) \log(dA_h^\theta(t_{ij}; \boldsymbol{X}_{ij})) & - & \sum_{j=1}^{n_i} (z_{i0} + z_{ij}) \Lambda_{ij} \quad + \\ \sum_{j=1}^{n_i} N_{ij}(\tau) \log(z_{i0} + z_{ij}) & + & \sum_{j=1}^{n_i} \log(p(z_{ij})). \end{array} \tag{22}$$

The first line of (22) contains the unknown intensity, and the E-step of the algorithm consists of calculating the conditional expectations of the frailties given the data, i.e.,

$$E\{Z_{ij} \mid (\boldsymbol{N}, \boldsymbol{Y})\} \; ; \; i = 1, \ldots, n; j = 0, \ldots, n_i.$$

The conditional distributions of the frailties given the data are not in general gamma. Rather, they are gamma mixtures with expectations that are readily calculated. For $Z_0 \mid (\boldsymbol{N}, \boldsymbol{Y})$, we find the following conditional expectation

$$E\{Z_0 \mid \boldsymbol{N}, \boldsymbol{Y}\} = \frac{1}{\Lambda_\cdot + \eta} \frac{\sum_{j=0}^m C_j^{(m)} \frac{\Gamma(\nu+j+1)}{(\Lambda_\cdot+\eta)^j}}{\sum_{j=0}^m C_j^{(m)} \frac{\Gamma(\nu+j)}{(\Lambda_\cdot+\eta)^j}}. \tag{23}$$

Only in the case of no deaths does the distribution of $Z_0 \mid (\boldsymbol{N}, \boldsymbol{Y})$ reduce to a Γ-distribution with parameters $(\nu, (\Lambda_\cdot + \eta)^{-1})$ and conditional expectation $\nu/(\Lambda_\cdot + \eta)$.

For a surviving individual indexed by j, $Z_j \mid (\boldsymbol{N}, \boldsymbol{Y})$, is gamma distributed with parameters $(\nu^*, (\Lambda_j + \eta)^{-1})$. This immediately yields the following conditional expectation

$$E\{Z_j \mid \boldsymbol{N}, \boldsymbol{Y}\} = \frac{\nu^*}{\Lambda_j + \eta}. \tag{24}$$

For a individual that dies, indexed by j, we find the following conditional expectation

$$E\{Z_j \mid \boldsymbol{N}, \boldsymbol{Y}\} = \frac{\nu^*}{\Lambda_j + \eta} \frac{\sum_{k=0}^{m-1} \frac{C_k^{(m-1)} \Gamma(\nu+k)}{(\Lambda_\cdot+\eta)^k} \left(\frac{\nu+k}{\Lambda_\cdot+\eta} + \frac{\nu^*+1}{\Lambda_j+\eta} \right)}{\sum_{k=0}^m C_k^{(m)} \frac{\Gamma(\nu+k)}{(\Lambda_\cdot+\eta)^k}}. \tag{25}$$

The M-step of the EM-algorithm, now consists of a maximization of (6) where all unobserved variables are replaced by their conditional expectations calculated in the E-step. The first line of (6), for known frailties, is of the form of the partial likelihood for a Cox regression model. Therefore, for a fixed value of the regression parameter $\boldsymbol{\beta}$, an estimate of the cumulative baseline hazard $A_h(t) = \int_0^t dA_h(s)$ is given by the Nelson-Aalen estimator

$$\widehat{A}_h(t) = \int_0^t \frac{dN_h(s)}{\sum_{H_h} (\hat{z}_{i0} + \hat{z}_{ij}) \exp(\boldsymbol{\beta}^T X_{ij}) Y_{ij}(s)} \tag{26}$$

where $N_h(t) = \sum_h N_{ij}(t)$ is the sum over all subjects in stratum h, and H_h is the set of subjects in stratum h, i.e. $H_h = \{(i,j) \mid h(i,j) = h\}$. Furthermore, an estimate of $\boldsymbol{\beta}$ is

given by maximization of the *Cox partial likelihood*

$$\prod_{i,j} \left(\frac{(\hat{z}_{i0} + \hat{z}_{ij}) \exp(\boldsymbol{\beta}^T X_{ij})}{S_{h(i,j)}(\boldsymbol{\beta})} \right)^{N_{ij}(\tau)} \tag{27}$$

where

$$S_h(\boldsymbol{\beta}) = \sum_{H_h} (\hat{z}_{i0} + \hat{z}_{ij}) \exp(\boldsymbol{\beta}^T X_{ij}) Y_{ij}(t_{ij}). \tag{28}$$

These estimates, i.e., $\widehat{\boldsymbol{\beta}}$ and $\widehat{A}_h(t)$, may also be found using standard software for Cox regression analysis. This is done by introducing the estimated frailties or rather $(\log(\hat{z}_{i0} + \hat{z}_{ij}))$ as covariates with a predetermined associated regression parameter of 1.

For the estimation of the frailty parameters (ν, ν^*, η) we suggest, in line with the approach taken in Section 3.2, a two-dimensional search procedure. The maximization procedure is then iterative where in each iteration, a full EM-algorithm is carried out followed by the two dimensional search. Also, variance estimates can be obtained as described in the one-dimensional case, though calculations become rather involved.

5 Further Developments

The above litter frailty model can be viewed as corresponding to a one-way analysis of variance for normally distributed variables. In principle, the analogy extends to a multi-way ANOVA which could be defined and calculations carried out along the same lines as shown in this paper. Also, instead of only one individual frailty (Z_{ij}), one could introduce a concept of dynamic frailties, the idea being that frailty changing over time captures, for example, the changing environmental exposures. This idea has been discussed by Paik et al. (1994).

Our current research is on models to analyse data from adoption studies. The data include life time events for a number of adoptive children, as well as for both their biological and adoptive relatives. The aim of these studies is to identify both a genetic and environmental influence on some event where the design ensures that these effects are not confounded; roughly, a genetic effect will induce an association between adoptee and biological family, whereas an environmental effect implies an association between adoptee and adoptive family. With a large family structure on both the biological and adoptive side, the correlation structure becomes complicated. As an illustration of these models, we mention the simplest possible adoption study design in which the environmental and genetic effects can be separated, one consisting of the adoptive mother (A), the adoptive child (C) and the biological mother (B). A similar kind of model involving (biological) mother, father and child, but focusing only on genetic effects is studied by Andersen and Korsgaard (1995). A natural additive structure accounting for the association between life

times due to genetic and environmental effects, would be

$$\begin{array}{rcccccccc} Z_A & = & G_M^A & + & G^A & + & E_s^A & + & E_{ns}^A \\ & & & & & & | & & \\ Z_C & = & G_M^B & + & G & + & E_s^A & + & E_{ns} \\ & & | & & & & & & \\ Z_B & = & G_M^B & + & G^B & + & E_s^B & + & E_{ns}^B \, . \end{array}$$

Here, each individual has for reasons of symmetry, four frailties, two interpreted as genetic (G) and two environmental (E) contributions. The genetic contributions for each individual come from that individuals biological mother and father, and the adoptive child therefore shares one of the two genetic frailties with its biological mother (G_M^B). Similarly, the adoptee shares (some) environment with the adoptive mother (E_s^A). The remaining frailties are specific to each individual. Now assume that all frailties are independent and gamma distributed, genetic frailties with parameters (ν_G, η), frailties associated with shared environment (subscript s) with parameters (ν_s, η) and frailties associated with non-shared environment (subscript ns) with parameters (ν_{ns}, η). As previously, we restrict the mean frailty to 1 by setting $\eta = \nu_G + \nu_s + \nu_{ns}$. Estimation and inference can now, in principle, be carried out as for the litter frailty model. However, the lack of symmetry results in very involved expressions, especially in the E-step of the algorithm. For larger family structures the number of distinct expressions needed grows at an exponential rate. There seems to be scope for further study here.

6 Discussion

In this paper, we have intended to show how frailty models can be used to model multivariate event time data, i.e., multiple events that are correlated. Whereas this originally has been done using the standard (one-dimensional) frailty model, we argue, following Yashin et al. (1995), that with an additive frailty structure, we avoid the possible errors due to confounding of heterogeneity or overdispersion within a population and the correlation between event times.

There are many potential applications of these models, some of them sketched for the simple litter frailty model in Section 4. Other applications include correlated competing risks, i.e., when more than one cause of death is considered, and where it is conceivable that an increased risk of dying of one cause is associated with an increased risk of dying of one or more other causes. As a Bayesian application of the models, they could be used to estimate individual yield in an agricultural context, breeding, say. The aim could be to identify animals whose offspring reach a certain target in a certain time. An insurance application in which individuals are insured against different events pose an alternative yield estimation problem. Suppose that each individual may experience different events e.g. accidents that require the insurance company to pay. From the company's perspective, it would be interesting to identify individuals with particularly high propensity to accidents. A possible model for this set-up, is a litter frailty model where the basic unit (the litter)

is each subject. Associated with each subject is a number of processes, one for each type of event (accidents). Each type of event has its own associated baseline hazard function corresponding to the strata of the model. The shared litter frailty, i.e., now the subject specific frailty, would represent (for the insurance company) a single measure of that person's propensity or frailty to get involved in accidents, and could form the basis for an individual premium policy.

We have not tried to give an exhaustive overview of the different approaches, e.g. based on generalised estimating equations (Liang et al., 1993) or different attempts to define multivariate survival distribution functions (Andersen et al., 1993, Chapter X.3). Rather we have based the model on the conditional (given the frailties) hazard function as a measure of individual risk. This is the relevant measure in many biostatistical applications. Within this framework of conditional models, we have demonstrated that the additive frailty structure is mathematically tractable. However, it is by no means the only way of modelling correlated event times data, and we point out two alternatives that have particular appeal. The first is conceptually closely connected with the litter model, only the frailty structure is assumed multiplicative $Z_i^{(j)} = Z_{i0}Z_{ij}$, as opposed to additive $Z_i^{(j)} = Z_{i0} + Z_{ij}$, cf. (12), i.e., each individual within litters is endowed with two independent frailties that act multiplicatively on the hazard. With this structure event times of individuals within litters are still correlated through their shared frailty. Furthermore, it is a more parsimonious model since each frailty can be interpreted as an uobserved covariate in a Cox regression model where each frailty (Z) enters the exponential regression function as $\log(Z)$. With this structure, the gamma frailty distribution is less attractive because it is not invariant to multiplication. An appealing choice would be the log-normal distribution because the logarithms of the frailties, interpreted as covariates in a Cox-regression, would the be normal. However, this model is not mathematically tractable in that it does not yield a closed form expression for the observed data likelihood leaving us with for example Markov Chain Monte Carlo methods, see Clayton (1991). This approach in connection with multiplicative models is discussed by Thomas (1995).

A second alternative which seems attractive is the model of Aalen (1980, 1989) where, for individual i, the hazard function

$$\alpha_i(t) = \beta_0(t) + \boldsymbol{\beta}(t)^T \boldsymbol{X}_i(t)$$

is additive in the covariates $\boldsymbol{X}_i$, and where $\beta_l(t), l = 0, \ldots, p$ are unknown (non-parametric) regression functions. In this model, an additive frailty structure would fit nicely with the postulated effect of the observed covariates. However, since the standard inference in this model is not likelihood based, it is less obvious how to incorporate frailties into this model. A model of this type involving parametric regression functions is discussed by Zahl (1994).

It is important to bear in mind that none of these models are based on a biological or genetic theory, and they should not be interpreted as such. Rather, we feel that they address the important question of correlated event times as well as heterogeneity, in a fairly straightforward way, through the specification of the conditional intensity given the frailties. That the heterogeneity in the real world is of a much more complicated nature

than can be captured by these relatively simple additive frailty models goes without saying. Yet, we feel that they provide an attempt to deal with a problem known to cause biased and inefficient estimates.

Acknowledgements. The research was partially supported by a grant (R01 CA54706-04) from the National Cancer Institute. Jørgen Holm Petersen and Per Kragh Andersen are supported by the Danish Epidemiology Science Centre; Jørgen Holm Petersen to the branch at the Institute of Preventive Medicine, Copenhagen Hospital Coporation and Per Kragh Andersen to the branch at Statens Seruminstitut, Copenhagen. The activities of the Danish Epidemiology Science Center are supported by a grant from the Danish National Research Foundation.

References

Aalen, O.O. (1980). A model for non-parametric regression analysis of counting processes. *Springer Lecture Notes in Statistics* **2**, 1–25.

Aalen, O.O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine* **8**, 907–925.

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

Andersen, A.H. and Korsgaard, I.R. (1995). The additive genetic gamma frailty model. Technical Report 315, Department of Theoretical Statistics, University of Aarhus, Denmark.

Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. (1993). *Statistical models based on counting processes.* Springer, New York.

Andersen, P.K., Klein, J.P., Knudsen, K., and Palacios, R.T. (1995). Estimation of variance in Cox's regression model with gamma frailties. Technical report, Department of Biostatistics, University of Copenhagen, Denmark.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer-Verlag, New York.

Bretagnolle, J. and Huber-Carol, C.. (1988). Effects of omitting covariates in Cox's model for survival data. *Scandinavian Journal of Statistics* **15**, 125–138.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151.

Clayton, D.G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* **47**, 467–485.

Clayton, D.G. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A* **148**, 82–117.

Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

Dempster, A.P., Laird, N.M., and Rubin, D.R. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* **39**, 1–38.

Elbers, C. and Ridder, G. (1982). True and spurious duration dependence: the identifiability of the proportional hazard model. *Review of Economic Studies* **XLIX**, 403–409.

Gill, R.D. (1985). Discussion of the paper by D. Clayton and J. Cuzick. *Journal of the Royal Statistical Society, Series A* **148**, 108–109.

Hougaard, P. (1995). Frailty models for survival data. To appear in *Lifetime Data Analysis*, vol. 1.

Iachine, I.A. (1995). Parameter estimation in the bivariate correlated frailty model with observed covariates via the EM-algorithm. Population Studies of Aging 16, Centre for Health and Social Policy, Odense University.

Klein, J.P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* **48**, 795–806.

Kortram, R.A., Lenstra, A.J., Ridder, G., van Rooij, A.C.M. (1995). Constructive estimation of the mixed proportional hazards model. *Statistica Neerlandica* **49**, 269–281.

Liang, K., Self, S.G., and Chang, Y. (1993). Modelling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society, Series B* **55**, 441–453.

Louis, T.A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society, Series B* **44**, 226–233.

Murphy, S.A. (1994). Consistency in a proportional hazards model incorporating a random effect. *The Annals of Statistics* **22**, 712–731.

Murphy, S.A. (1995). Asymptotic theory for the frailty model. *The Annals of Statistics* **23**, 182–198.

Nielsen, G.G., Gill, R.D., Andersen, P.K., and Sørensen, T.I.A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19**, 25–43.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association* **84**, 487–493.

Paik, M.C., Tsai, W., and Ottman, R. (1994). Multivariate survival analysis using piecewise gamma frailty. *Biometrics* **50**, 975–988.

Petersen, J.H. (1995). A litter frailty model. Technical report, Department of Biostatistics, University of Copenhagen, Denmark.

Self, S.G. and Prentice, R.L. (1986). Incorporating random effects into multivariate relative risk regression models. In Moolgavkar, S.H. and Prentice, R.L., editors, *Modern Statistical Methods in Chronic Disease Epidemiology*, 167–177. John Wiley and Sons, New York.

Thomas, D.C. (1995). A nonparametric individual frailty model for family survival data. Unpublished manuscript, University of Southern California.

Vaupel, J.W., Harvald, B., Holm, N.V., Yashin, A.I., and Xiu, L. (1992). Survival analysis in genetics: Danish twin data applied to a gerontological question. In Klein, J.P. and Goel, P.K., editors, *Survival Analysis: State of the Art*, 121–138. Kluwer Academic Publishers, New York.

Vaupel, J.W., Manton, K.G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.

Vaupel, J.W. and Yashin, A. (1985b). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *The American Statistician* **39**, 176–185.

Yashin, A., Vaupel, J., and Iachine, I. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. To appear in *Mathematical Population Studies*.

Zahl, P. (1994). Frailty modelling for the excess hazard. Technical report, Section of Medical Statistics, University of Oslo, Norway.