Estimation of variance in Cox's regression model with gamma frailties.

Per Kragh Andersen[1,2] John P. Klein[3] Kim M. Knudsen[2]
René Tabanera y Palacios[4]

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

[2] Danish Epidemiology Science Center
[3] Division of Biostatistics, Medical College of Wisconsin
[4] Novo Nordisk A/S Gentofte

September 1, 1995

Abstract

The Cox regression model with a frailty factor allows for unobserved heterogeneity or for statistical dependence between the observed survival times. Estimation in this model is reviewed and we address the problem of obtaining variance estimates for regression coefficients, frailty parameter, and cumulative baseline hazards using the observed non-parametric information matrix. A number of examples are given comparing this approach with fully parametric inference in models with piecewise constant baseline hazards.

Key words: frailty model; survival analysis; EM algorithm; non-parametric MLE; observed information.

1 Introduction.

In the analysis of survival data or more general event history data the assumption is frequently made that the life histories for the individuals under study are all statistically independent (at least conditionally on observed time-fixed covariates). One alternative model which does allow for dependence between related individuals is the shared frailty model which has been studied by a number of authors over the past years, including Clayton (1978, 1991), Clayton and Cuzick (1985), Hougaard (1986), Nielsen et al. (1992), Klein (1992), and Morsing (1994). Another application of the frailty model is to interpret the frailty as modeling the effects of unobserved covariates (e.g., Vaupel et al., 1979; Hougaard, 1984; and Murphy, 1994, 1995). There are, however, difficulties in including both dependence and heterogeneity in the shared frailty model (Yashin et al., 1995). We return to this in Section 4.

The purpose of the present paper is, first, to address the problem of assessing the uncertainty of the parameter estimates from the semi-parametric frailty model and, second, to present a number of examples using this model and some of its fully parametric counterparts. The structure of the paper is as follows: In Section 2, the shared frailty model is presented and inference in the model is discussed with special emphasis on how to obtain variance estimates for the estimated parameters. (Some of the technical details are deferred to an Appendix.) Section 3 contains a series of examples using the frailty model while, in Section 4, we discuss our results and give some concluding remarks.

2 Inference in the shared gamma-frailty model.

Consider independent *groups* (e.g., families or litters) of individuals indexed by $i = 1, \cdots, n$, group i consisting of $n_i \geq 1$ *individuals* indexed by $l = 1, \cdots, n_i$. For each (i, l), a multivariate counting process $(N_{hil}(t), h = 1, \cdots, k)$ is observed, index h referring to either different *types* of events which an individual may experience during his or her life, or to different *strata* to which (i, l) may belong, see, e.g., Andersen et al.(1993, Section VII.1) for examples. For each (h, i, l) we further observe a process $Y_{hil}(t)$ indicating whether or not individual (i, l) is observed to be at risk for experiencing an event of type h at time $t-$ (or whether or not individual (i, l) is observed to belong to stratum h at time $t-$). Finally, for each (h, i, l) (possibly time-dependent) covariates $\boldsymbol{X}_{hil}(t)$ are observed.

A model for $N_{hil}(t)$ is now set up via its intensity process $\lambda_{hil}(t)$ as follows. For given values of independent, group specific *frailties* $z_i, i = 1, \cdots, n$, the intensity process is given by

$$\lambda_{hil}(t \mid \boldsymbol{z}) = z_i Y_{hil}(t) \alpha_{h0}(t, \boldsymbol{\gamma}) \exp(\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{X}_{hil}(t)), \tag{1}$$

i.e., individuals $l = 1, \cdots, n_i$ in group i share the value, z_i, of the unobserved frailty. Here $\boldsymbol{\beta} = (\beta_1, \cdots \beta_p)^{\mathsf{T}}$ is a vector of unknown *regression coefficients* describing the effect of the observable covariates $\boldsymbol{X}_{hil}(t)$, and $\alpha_{h0}(t, \boldsymbol{\gamma})$ is a *baseline hazard* for type or stratum $h, h = 1, \cdots, k$, parametrized by unknown parameters $\boldsymbol{\gamma}$. A finite dimensional $\boldsymbol{\gamma} \in \boldsymbol{R}^q$ corresponds to a fully parametric model while a semi-parametric model with completely unspecified baseline hazards $\alpha_{h0}(t)$ may be obtained for an infinite dimensional $\boldsymbol{\gamma}$. Examples of parametric models include a Weibull baseline: $\alpha_0(t) = \alpha \rho t^{\rho - 1}$ where $\boldsymbol{\gamma} = (\alpha, \rho)$ and a piecewise constant baseline: $\alpha_0(t) = \gamma_k$ when $t \in [t_{k-1}, t_k\ ), k = 1, \ldots, q$, where $0 = t_0 < t_1 < \ldots < t_q = \tau \leq +\infty$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots \gamma_q)$. The latter model (without the frailty factor) is a version of a *Poisson* regression model (e.g., Clayton and Hills, 1993).

If the frailties $\boldsymbol{z}$ were observed, the (partial) likelihood for $(\boldsymbol{\gamma}, \boldsymbol{\beta})$ would be

$$L(\boldsymbol{\gamma}, \boldsymbol{\beta} \mid \boldsymbol{z}) = \prod_i \left(\prod_{h,l} \exp(-\int_0^\tau \lambda_{hil}(u \mid \boldsymbol{z}) \mathrm{d}u) \prod_t \lambda_{hil}(t \mid \boldsymbol{z})^{\mathrm{d}N_{hil}(t)} \right), \quad (2)$$

$\tau \leq +\infty$ denoting the upper limit for the observation times. Now, $\boldsymbol{z}$ is not observed, so (2) cannot be computed from the observations. Assuming $z_i, i = 1, \cdots, n$ to be *i.i.d* with density $f(z, \boldsymbol{\theta})$ and assuming censoring to be non-informative on the frailties $\boldsymbol{z}$ (Nielsen et al., 1992) the marginal (partial) likelihood for the observed data is

$$\begin{aligned} L(\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \prod_i \int_0^\infty \left(\prod_{hl} \exp(-\int_0^\tau \lambda_{hil}(u \mid \boldsymbol{z}) \mathrm{d}u) \right. \\ &\times \left. \prod_t \lambda_{hil}(t \mid \boldsymbol{z})^{\mathrm{d}N_{hil}(t)} \right) f(z_i, \boldsymbol{\theta}) \mathrm{d}z_i. \end{aligned} \quad (3)$$

This likelihood will, obviously, depend on the assumed model $f(z, \boldsymbol{\theta})$ for the frailties. In what follows, we shall make the convenient assumption (e.g., Clayton, 1978; Vaupel et al., 1979; Nielsen et al., 1992; Klein, 1992) that z_i is gamma-distributed with mean 1 and variance θ, i.e.,

$$f(z, \theta) = \frac{z^{\frac{1}{\theta}-1} e^{-\frac{z}{\theta}}}{, (\frac{1}{\theta}) \theta^{\frac{1}{\theta}}}$$

but it should be emphasized that, in principle, the following derivations can be carried out for other choices of frailty distribution, see, e.g., Klein et al. (1992). In the gamma-model, the likelihood (3) reduces to

$$L(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta) = \prod_i \frac{, (N_i + \frac{1}{\theta})(\frac{1}{\theta})^{\frac{1}{\theta}}}{, (\frac{1}{\theta})(\Lambda_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) + \frac{1}{\theta})^{N_i + \frac{1}{\theta}}} \prod_{hlt} \left(\alpha_{h0}(t, \boldsymbol{\gamma}) \exp(\boldsymbol{\beta}^\mathsf{T} \boldsymbol{X}_{hil}(t)) \right)^{\mathrm{d}N_{hil}(t)}, \quad (4)$$

where

$$N_i = \sum_{h=1}^{k} \sum_{l=1}^{n_i} N_{hil}(\tau)$$

is the total number of events in group i, and $\Lambda_i(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is given by

$$\Lambda_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{h=1}^{k} \int_0^\tau S_{hi}^{(0)}(\boldsymbol{\beta}, u) \alpha_{h0}(u, \boldsymbol{\gamma}) \mathrm{d}u, \quad (5)$$

with

$$S_{hi}^{(0)}(\boldsymbol{\beta}, u) = \sum_{l=1}^{n_i} Y_{hil}(u) \exp(\boldsymbol{\beta}^\mathsf{T} \boldsymbol{X}_{hil}(u))$$

being the sum over the type h risk set at time $u-$ for group i.

In a parametric model, estimates for $(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta)$ are obtained by maximising (4) either directly or by using the EM-algorithm. Variance estimates may then be based on the inverse information matrix $\boldsymbol{I}(\boldsymbol{\eta})^{-1}$, where $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_{q+p+1}) = (\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta)$ and

$$\boldsymbol{I}(\boldsymbol{\eta}) = \frac{\partial^2}{\partial \boldsymbol{\eta}^2} l(\boldsymbol{\gamma}, \boldsymbol{\beta}, \theta)$$

with $l(\cdot) = -\log L(\cdot)$. A version of a model with a piecewise constant baseline hazard was studied by Thall (1988) and more recently by Knudsen (1994). For

a semi-parametric model it was discussed by Nielsen et al. (1992), Klein (1992), and Andersen et al. (1993, Chapter IX) (all based on Gill, 1985) how to use the EM-algorithm for maximising (4). It is easily seen that a cumulative baseline hazard

$$A_{h0}(t) = \int_0^t \alpha_{h0}(u)\mathrm{d}u, h = 1, \cdots, k,$$

maximising (4) will only put mass on the observed type h event times, say, $T_{hj}, j = 1, \cdots, m_h$, that is, we can write the estimator as

$$\widehat{A_{h0}(t)} = \sum_{j:T_{hj} \le t} \widehat{\alpha_{hj}}. \tag{6}$$

We write the parameter

$$A_{h0}(t) = \sum_{j:T_{hj} \le t} \alpha_{hj}$$

similarly for some unknown jumps $\alpha_{hj}, j = 1, \cdots, m_h$, for the cumulative baseline hazard $A_{h0}(\cdot)$ (here m_h is the number of distinct type h event times, i.e., $m_h = \sum_i \sum_l N_{hil}(\tau)$ only if there are no ties among the type h event times) and the $\Lambda_i(\boldsymbol{\beta}, \boldsymbol{\gamma})$ given by (5) become

$$\Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{h=1}^{k} \sum_{j=1}^{m_h} S_{hi}^{(0)}(\boldsymbol{\beta}, T_{hj})\alpha_{hj}$$

where $\boldsymbol{\alpha} = ((\alpha_{hj}, j = 1, \cdots, m_h), h = 1, \cdots, k)$. Thus, one needs to maximise (4) over the $m + p + 1$ parameters $\boldsymbol{\eta} = (\boldsymbol{\alpha}, \beta_1, \cdots, \beta_p; \theta)$ where $m = \sum_h m_h$. Minus the log-likelihood for these parameters are

$$\begin{aligned} l(\boldsymbol{\eta}) &= \sum_{i=1}^{n} \Bigg(- \sum_{r=0}^{N_i - 1} \log(r + \frac{1}{\theta}) \\ &+ \frac{1}{\theta} \log \theta + (N_i + \frac{1}{\theta}) \log(\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})) \\ &- \sum_{h=1}^{k} \sum_{l=1}^{n_i} \sum_{j=1}^{m_h} \mathrm{d}N_{hil}(T_{hj}) \boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{X}_{hil}(T_{hj}) \Bigg) - \sum_{h=1}^{k} \sum_{j=1}^{m_h} m_{hj} \log \alpha_{hj}, \end{aligned} \tag{7}$$

where $m_{hj} = \sum_{i=1}^{n} \sum_{l=1}^{n_i} \mathrm{d}N_{hil}(T_{hj})$ is the number of type h events at time T_{hj}. It is easily seen that for model with a model with a piecewise constant baseline hazard with the *smallest possible interval length* (e.g., 1 day if the survival times are in days) the likelihood is also given by (7). To obtain variance estimates for the parameter estimates in the semi-parametric case we suggest using the same procedure as for the parametric model, i.e., to calculate the inverse information matrix $\boldsymbol{I}(\boldsymbol{\eta})^{-1}$, where

$$\boldsymbol{I}(\boldsymbol{\eta}) = \frac{\partial^2}{\partial \boldsymbol{\eta}^2} l(\boldsymbol{\eta}).$$

Since $\widehat{A_{h0}(t)}$ is a linear function of the $\widehat{\alpha_{hj}}$, see (6), estimation of the variance of $\widehat{A_{h0}(t)}$ is straightforward. In the Appendix we present the formulas for the elements of $\boldsymbol{I}(\boldsymbol{\eta})$. Obviously, since the number, $m+p+1$, of parameters increases with the sample size this is a non-standard approach and there is no theory ensuring that consistent variance estimators are obtained in this way. However, Murphy (1995) studied large sample properties in a special case ($k = 1$: one stratum; $p = 0$: no covariates) of our model (1) finding conditions under which $\widehat{\theta}$

and $\widehat{A_0(\cdot)}$ have an asymptotic joint normal distribution and deriving a consistent variance estimator which is asymptotically equivalent to ours. Furthermore, in the next section we present a number of examples comparing results obtained with this approach with results from approximating parametric models with a piecewise constant baseline hazard.

It should be noted that Klein (1992) used a related method for finding variance estimates for $(\widehat{\boldsymbol{\beta}}, \widehat{\theta})$ based on inserting the value

$$\widehat{\alpha_{hj}} = \frac{m_{hj}}{\sum_{i=1}^{n} \widehat{z}_i S_{hi}^{(0)}(\boldsymbol{\beta}, T_{hj})}$$

into (7) and treating the resulting function of $(\boldsymbol{\beta}, \theta)$ as a profile likelihood. Here, $\widehat{z}_i$ is the predicted frailty for group i obtained in the E-step of the EM-algorithm and $\widehat{\alpha_{hj}}$ is the value obtained in the M-step. However, by neglecting that

$$\widehat{z}_i = \frac{\frac{1}{\theta} + N_i}{\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})}$$

depends on all the parameters this procedure is likely to produce variance estimates which are too low. We shall study this in more detail in the examples in the next section.

3 Examples.

To illustrate these techniques we present three examples. The first two examples were previously analysed by Nielsen et al. (1992) while the final example, drawn from the Framingham heart study (Dawber, 1980), was analysed by Klein et al. (1992). For each example we have a single stratum and the processes $Y_i(t)$ are indicators for individual i being at risk just prior to time t, i.e., in Examples 1 and 2 $Y_i(t) = I(\widetilde{X}_i \geq t)$ where $\widetilde{X}_i$ is the observation time and in Example 3, $I(V_i < t \leq \widetilde{X}_i)$ where V_i is the time of entry into the study of individual i. For each example we fitted a semi-parametric Cox regression model and models with piecewise constant baseline hazards. In Example 1 we, furthermore, fitted a model with a Weibull baseline hazard. The regression and frailty parameters and their associated standard errors were estimated as described in the previous section. For comparison for the semiparametric approach we include the estimated standard error of the parameter estimates based on the formulation given by Klein (1992).

Example 1.

The data used in this example, published by Mantel et al. (1977), come from a litter-matched tumorigenesis experiment with one drug treated and two placebo treated rats per litter. One might expect that times to tumor formation for rats in a given litter are correlated due to genetic or shared environmental effects. Times are given in weeks, and death before tumor occurrence yield right-censored observations. Table 1 presents the results.

	Treatment		Frailty	
Model	$\widehat{\beta}$	(S.D.)	$\widehat{\theta}$	(S.D.)
Cox model (frailty)	0.904	(0.323)(0.323)*	0.472	(0.462)(0.431)*
Cox model (independence)	0.897	(0.317)	0	(-)
Weibull (frailty)	0.908	(0.322)	0.492	(0.470)
Weibull (independence)	0.905	(0.380)	0	(-)
Piecewise constant (frailty, $q = 6$)	0.907	(0.323)	0.485	(0.468)
Piecewise constant (independence, $q = 6$)	0.907	(0.317)	0	(-)
Piecewise constant (frailty, $q = 31$)	0.904	(0.323)	0.472	(0.462)
Piecewise constant (independence, $q = 31$)	0.897	(0.317)	0	(-)

Table 1. Results from analyses of the data of Mantel et al. (1977). Asterisks indicate standard errors obtained following Klein (1992).

The results show good agreement between the various parametric and semi-parametric models. As expected, the standard errors from Klein's (1992) formulation are (slightly) smaller than those found using the results in this report. The close agreement between the estimates based on the Cox regression model and the models with a piecewise constant hazard provide some evidence that our suggested variance estimator is reasonable. It is well known that (in the absence of a frailty factor) results from Cox and Poisson regression models give results that tend to be very similar and in this example we have seen that the same is true for the gamma frailty model: with $q = 6$ intervals (each of length 13 weeks) the results are close and with $q = 31$ intervals (equal to the number of different failure times in the data set) our results from the two models are identical (which they should be since the likelihoods as mentioned above are identical). Also the likelihood ratio tests for independence gave comparable results: 1.52 for the Cox model 1.62 for the Weibull, and 1.58 and 1.52 for the two models with piecewise constant hazards. In this litter-matched study an obvious alternative approach would be to use a standard *stratified* Cox regression model, i.e., a model with a separate baseline hazard for each litter. In this analysis the estimated treatment effect is $\widehat{\beta} = 0.880$ with a somewhat higher standard deviation (0.377) than seen in Table 1.

Example 2.

The second example is based on data reported by Batchelor and Hackett (1970) on sixteen severely burned patients treated with skin allografts. Patients received one to four skin allografts from donors who were either closely or poorly HL-A tissue matched with the patients. The event of interest is the time (in days) to rejection of the graft due to an immune response by the patient and, since the immune response is likely to depend on factors like the patient's genetic constitution and preceeding stimulation, times within a patient may be correlated. Here, the models are fitted with a single covariate indicating close or poor HL-A matching. Since the Weibull model in this example did not provide a reasonable fit under the independence assumption only results from Cox models and models with piecewise constant hazards are presented in Table 2.

	close match		Frailty	
Model	$\widehat{\beta}$	(S.D.)	$\widehat{\theta}$	(S.D.)
Cox model (frailty)	-1.166	(0.484)(0.468)*	0.555	(0.580)(0.335)*
Cox model (independence)	-1.029	(0.426)	0	(-)
Piecewise constant (frailty, $q = 6$)	-1.177	(0.487)	0.652	(0.503)
Piecewise constant (independence, $q = 6$)	-1.104	(0.448)	0	(-)

Table 2. Results from analyses of the data of Batchelor and Hackett (1970). Asterisks indicate standard errors obtained following Klein (1992).

In this example, the Klein (1992) estimates for the standard errors in the Cox model with frailty are quite a bit smaller than those based on the methods described in Section 2 above. Furthermore, the estimated frailty parameter using $q = 6$ intervals (each of length 2 weeks) is slightly larger than for the Cox model (which is this case corresponds to $q = 17$ intervals). The likelihood ratio tests for independence are 1.34 for the Cox model and 1.42 for the gamma Poisson model with $q = 6$.

Example 3.

As a third example we consider a cohort of 1571 individuals selected from the Framingham Heart Study, see Dawber (1980) for details. Subjects were included in the sample if they reached age 45 with no prior history of hypertension or glucose intolerance while in the Framingham Study prior to age 45. Covariates included in the model were body mass index (BMI) in kg/m^2, cholesterol (CHOL) in mg/dL, sex (males=1, females=0), smoking status (smokers=1), and hypertension status (HYP) (hypertensive or borderline hypertensive =1). The covariate used for each subject was the value of the variate measured at the examination closest to age 45 and the entry time used in our analyses is the age at this examination. Since the patients enter the Framingham Study at different ages it is possible that there may be a cohort effect on the outcome and to test this we included the waiting time on study to inclusion as a covariate. This was, however, found to be insignificant in all models and is excluded from the results presented here.

The endpoint of interest is the first evidence of coronary heart disease (CHD) which occurred in 250 individuals. In this example, siblings, who share a common genetic code and a common environment in childhood, have a common value of the frailty. In this example, we looked at the Cox model and models with piecewise constant baseline hazards with and without frailty, see Table 3.

			Frailty models			
Model	Cox model		piecewise constant ($q = 6$)		piecewise constant ($q = 27$)	
Variable	$\widehat{\beta}$	(S.D.)	$\widehat{\beta}$	(S.D.)	$\widehat{\beta}$	(S.D.)
BMI	0.373	(0.204)(0.203)*	0.366	(0.197)	0.372	(0.205)
CHOL	0.447	(0.169)(0.169)*	0.439	(0.165)	0.450	(0.170)
SMOKE	0.352	(0.160)(0.159)*	0.333	(0.155)	0.353	(0.161)
SEX	-0.748	(0.162)(0.152)*	-0.716	(0.154)	-0.756	(0.164)
HYP	0.393	(0.163)(0.161)*	0.375	(0.157)	0.395	(0.164)
Frailty	0.835	(0.604)(0.375)*	0.562	(0.503)	0.894	(0.623)
			Independence models			
Model	Cox model		piecewise constant ($q = 6$)		piecewise constant ($q = 27$)	
Variable	$\widehat{\beta}$	(S.D.)	$\widehat{\beta}$	(S.D.)	$\widehat{\beta}$	(S.D.)
BMI	0.360	(0.182)	0.359	(0.183)	0.359	(0.182)
CHOL	0.426	(0.152)	0.423	(0.153)	0.427	(0.152)
SMOKE	0.318	(0.145)	0.312	(0.145)	0.317	(0.146)
SEX	-0.651	(0.135)	-0.652	(0.135)	-0.652	(0.135)
HYP	0.351	(0.143)	0.348	(0.144)	0.351	(0.144)

Table 3. Results from analyses of the Framingham data. Asterisks indicate standard errors obtained following Klein (1992).

Again in this example, a nice agreement is seen between the Cox models and the parametric models with large numbers of intervals (here, $q = 27$ corresponding to 1-year intervals) both with repect to the estimates and to their associated standard errors. For $q = 6$ (5-year intervals) both the frailty parameter and its S.D. is smaller than for $q = 27$ similar to what we saw in Example 2. Though, in

this example times are given with an accuracy of 1 month, we did not go beyond $q = 27$ since this would have required a substantial computer storage capacity. Comparing the standard errors based on the non-parametric information matrix with those based on Klein (1992) we see that the latter are slightly smaller for $\widehat{\beta_1}, \cdots, \widehat{\beta_5}$ but substantially smaller for $\widehat{\theta}$ - the same tendencies were present in Examples 1 and 2 but less pronounced. Comparing, finally, the models without frailties it is seen that both the estimates and their estimated standard errors are smaller in the frailty-less models. However, except for the effect of sex which was expected to change when a possible association within sibling pairs is accounted for, the differences here are rather small and the likelihood ratio tests for independence (3.01 for the Cox model, 1.75 for $q = 6$, and 3.30 for $q = 27$) are not significant. In this example, however, the Wald test statistic $(0.835/0.375)^2$ based on the standard error obtained following Klein (1992) corresponds to a P-value of 0.01.

4 Discussion.

In this paper, we have discussed how to use information calculations in a non-parametric setting: the Cox regression model with gamma frailties. If one uses the same procedure for the standard Cox regression model without frailty then it is quite easily seen that the inverse information matrix gives the variance estimates usually used for that model and which are known to be consistent (Andersen et al., 1993, Section VII.2).

Asymptotic results for the gamma frailty model without covariates were proved recently by Murphy (1995). Her results, however, do not cover the regression model that we have been studying in this paper but it is likely that her methods of proof may be adapted to this situation.

Morsing (1994) studied the performance of our variance estimator in a simulation study designed to investigate models for cross-over designs. He found the estimator to be in satisfactory agreement with the empirical variance of the parameter estimates over repeated samples.

Finally, since in our examples in Section 3 the variance estimates for the semi-parametric model based on the inverse information matrix came out very close to variance estimates in an approximating parametric model with piecewise constant baseline hazard there is some indication that our procedure may work in general.

On the other hand, this argument may be turned around to the conclusion that, at least in moderate and large samples, one may just as well use the Poisson regression model with frailties in situations where dependence or overdispersion is likely to be present in event history data.

Some care must be exercised, however, when one wishes to model simultaneously correlation and heterogeneity (Yashin, et al., 1995) since in the shared gamma frailty model there is only one parameter, θ, to account for both. These authors then suggested (for, e.g., twin studies) a correlated frailty model where the frailty for individual l in pair i is a sum of a pair specific term Z_{i0} and an individual specific term Z_{il} which are both assumed to follow a gamma distribution and which are independent. Inference in a slightly extended version of that model (using the EM algorithm) was discussed by Petersen et al. (1995) and information calculations as described in the present paper should also be possible in the correlated frailty model.

Acknowledgements. The research was partially supported by a grant (R01 CA54706-04) from the National Cancer Institute. The activities of the Danish

Epidemiology Science Center are supported by a grant from the Danish National Research Foundation.

5 Appendix: The information matrix based on (7).

To derive the information matrix from $l(\boldsymbol{\eta})$ given by (7) where

$$\boldsymbol{\eta} = ((\alpha_{hj}, j = 1, \cdots, m_h), h = 1, \cdots, k; \beta_1, \cdots, \beta_p; \theta)$$

we first obtain (minus) the score $\frac{\partial}{\partial \boldsymbol{\eta}} l(\boldsymbol{\eta})$:

$$\frac{\partial l}{\partial \alpha_{hj}} = \sum_{i=1}^{n} (N_i + \frac{1}{\theta}) \frac{S_{hi}^{(0)}(\boldsymbol{\beta}, T_{hj})}{\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})} - \frac{m_{hj}}{\alpha_{hj}},$$

$$\frac{\partial l}{\partial \beta_\nu} = \sum_{i=1}^{n} (N_i + \frac{1}{\theta}) \frac{\sum_{h=1}^{k} \sum_{j=1}^{m_h} S_{hi\nu}^{(1)}(\boldsymbol{\beta}, T_{hj}) \alpha_{hj}}{\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})} - \sum_{i=1}^{n} \sum_{h=1}^{k} \sum_{l=1}^{n_i} \sum_{j=1}^{m_h} \mathrm{d}N_{hil}(T_{hj}) X_{hil\nu}(T_{hj}),$$

$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^{n} \left(\sum_{r=0}^{N_i - 1} \frac{1}{r\theta^2 + \theta} + \frac{1}{\theta^2}(1 - \log \theta) - \frac{1}{\theta^2} \log(\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})) - \frac{N_i + \frac{1}{\theta}}{\theta + \theta^2 \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})} \right).$$

Here,

$$S_{hi\nu}^{(1)}(\boldsymbol{\beta}, t) = \sum_{l=1}^{n_i} Y_{hil}(t) X_{hil\nu}(t) \exp(\boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{X}_{hil}(t))$$

is the derivative of $S_{hi}^{(0)}(\boldsymbol{\beta}, t)$ with respect to β_ν. The second order derivatives are then:

$$\frac{\partial^2 l}{\partial \alpha_{hj} \partial \alpha_{h'j'}} = \sum_{i=1}^{n} -(N_i + \frac{1}{\theta}) \frac{S_{hi}^{(0)}(\boldsymbol{\beta}, T_{hj}) S_{h'i}^{(0)}(\boldsymbol{\beta}, T_{h'j'})}{(\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}))^2} + \delta_{(h,j),(h',j')} \frac{m_{hj}}{(\alpha_{hj})^2}.$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \alpha_{hj} \partial \beta_\nu} &= \sum_{i=1}^{n} \frac{N_i + \frac{1}{\theta}}{(\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}))^2} \left((\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})) S_{hi\nu}^{(1)}(\boldsymbol{\beta}, T_{hj}) \right. \\ &- \left. S_{hi}^{(0)}(\boldsymbol{\beta}, T_{hj}) \sum_{h=1}^{k} \sum_{j=1}^{m_h} S_{hi\nu}^{(1)}(\boldsymbol{\beta}, T_{hj}) \alpha_{hj} \right). \end{aligned}$$

$$\frac{\partial^2 l}{\partial \alpha_{hj} \partial \theta} = \sum_{i=1}^{n} S_{hi}^{(0)}(\boldsymbol{\beta}, T_{hj}) \frac{N_i - \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}))}{(1 + \theta \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}))^2}.$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_\nu \partial \beta_\mu} &= \sum_{i=1}^{n} \frac{N_i + \frac{1}{\theta}}{(\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}))^2} \\ &\times \left((\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})) \sum_{h=1}^{k} \sum_{j=1}^{m_h} S_{hi\nu\mu}^{(2)}(\boldsymbol{\beta}, T_{hj}) \alpha_{hj} \right. \\ &\left. -(\sum_{h=1}^{k} \sum_{j=1}^{m_h} S_{hi\nu}^{(1)}(\boldsymbol{\beta}, T_{hj}) \alpha_{hj})(\sum_{h=1}^{k} \sum_{j=1}^{m_h} S_{hi\mu}^{(1)}(\boldsymbol{\beta}, T_{hj}) \alpha_{hj}) \right), \end{aligned}$$

where $S^{(2)}_{hi\nu\,\mu}(\boldsymbol{\beta}, t)$ is the derivative of $S^{(1)}_{hi\nu}(\boldsymbol{\beta}, t)$ with respect to β_μ. Furthermore,

$$\frac{\partial^2 l}{\partial \beta_\nu \partial \theta} = \sum_{i=1}^{n} \frac{(N_i - \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}))(\sum_{h=1}^{k} \sum_{j=1}^{m_h} S^{(1)}_{hi\nu}(\boldsymbol{\beta}, T_{hj})\alpha_{hj})}{(1 + \theta\Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}))^2},$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \theta^2} &= \sum_{i=1}^{n} \left(\sum_{r=0}^{N_i - 1} (-\frac{1 + 2r\theta}{(r\theta^2 + \theta)^2}) \right. \\ &\quad \left. + \frac{1}{\theta^3}(2\log\theta - 3) + \frac{2}{\theta^3}\log(\frac{1}{\theta} + \Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})) + \frac{\frac{3}{\theta} + N_i + 2\Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha})(2 + \theta N_i)}{(\theta + \theta^2\,\Lambda_i(\boldsymbol{\beta}, \boldsymbol{\alpha}))^2} \right).\end{aligned}$$

References

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

Batchelor, J.R. and Hackett, M. (1970). HLA matching in treatment of burned patients with skin allografts. *Lancet* **2**, 581–583.

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). *J. Roy. Statist. Soc. A* **148**, 82–117.

Clayton, D.G. and Hills, M. (1993). *Statistical Models in Epidemiology.* Oxford University Press, Oxford.

Dawber, T.R. (1980). *The Framingham Study: The Epidemiology of Atherosclerotic Disease.* Harvard University Press, Cambrigde, Mass.

Gill, R.D. (1985). Discussion of the paper by D. Clayton and J. Cuzick. *J. Roy. Statist. Soc. A* **148**, 108–109.

Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71**, 75–83.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika* **73**, 671–678.

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

Klein, J.P., Moeschberger, M.L., Li, Y.H., and Wang, S.T. (1992). Estimating random effects in the Framingham heart study. In Klein, J.P. and Goel, P.K., editors, *Survival Analysis: Stat of the Art.* Kluwer, Dordrecht.

Knudsen, K.M. (1994). *Analysis of event history data in measles and diarrhea epidemiology.* PhD thesis, University of Copenhagen, Dept. of Biostatistics.

Mantel, N., Bohidar, N.R., and Ciminera, J.L. (1977). Mantel-Haenszel analyses of litter-matched time-to-response data, with modifications for recovery of interlitter information. *Cancer Res.* **37**, 3863–3868.

Morsing, T. (1994). Competing risks in cross-over designs. Preprint 20, Department of Mathematics, Chalmers Univ. of Technology, Göteborg.

Murphy, S.A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–731.

Murphy, S.A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **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. *Scand. J. Statist.* **19**, 25–43.

Petersen, J.H., Andersen, P.K., and Gill, R.D. (1995). Variance components models for survival data. Research report, Department of Biostatistics, Univ. Copenhagen.

Thall, P.F. (1988). Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics* **44**, 197–209.

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.

Yashin, A., Vaupel, J.W., and Iachine, I. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Math. Pop. Studies* . (in press).