**Focussed Model Selection for Frailty Models**

Gilbert MacKenzie[1]

[1] Centre of Biostatistics, Department of Mathematics & Statistics, The University of Limerick, Limerick, Ireland

## Abstract

In survival analysis recurrent event times are often observed on the same subject. These event times may be correlated and Cox's (1972) proportional hazards (PH) model has been extended by the inclusion of frailty (or random effect) terms to model this correlation. Typically, in such a model we deal with three unknown structures in the conditional hazard

$$\lambda_{ij}(t; x_i, u_i) = \lambda_0(t)\exp(x'_{ij}\beta).u_i \qquad (1)$$

where $i = 1, \ldots, m$ independent subjects and $j = 1, \ldots, n_i$ indexes the recurrent survival times on the $i$th subject whence $n = \Sigma n_i$. Formally, $\lambda_0(t)$ is an unknown function of possibly large dimension, $\beta$ is a regression parameter of fixed dimension $p$ and we may choose $u_i = \exp(z_i\nu)$ where $z_i$ is the $i$th row of a partitioned model matrix, $\nu = (\nu^1; \ldots; \nu^k)^T$ is a partitioned vector of $k$ distinct vector frailty components, each following a different distribution eg $\nu^r \sim N_{q_r}(0; \Sigma_r), r = 1, \ldots, k$. We assume there are maximally $d$ frailty dispersion parameters depending on how the $\Sigma_r$ are parametrized in terms of a vector $\alpha' = (\alpha^1, \ldots, \alpha^k)$, where dim $(\alpha) = d$.

With these arrangements, our interest lies in developing model selection criteria to identify the frailty structure best supported by the data. From this perspective we are inclined to view $\lambda_0(t)$ as a nuisance function and $\beta$ as a nuisance parameter, ie, classically we should like to have an inference function focussed on $\alpha$ alone, from which the nuisance components have been eliminated. We pursue this goal [1] within the framework of Hierarchical Generalized Linear Models using an *h*-likelihood approach (Lee & Nelder, 1996, 2001)

Accordingly, we develop an AIC for selecting the dispersion parameters which define the frailty structure by generalizing two HGLM information criteria to frailty models. Compared with classical random-effect models, inference about semi-parametric frailty models is complicated by censoring and by the presence of a nonparametric baseline hazard function. In particular, in semi-parametric frailty models the number of nuisance parameters in the baseline hazard function increases with sample size. The extension of the two HGLM criteria to frailty models can be developed naturally via a profile likelihood after eliminating the nuisance parameters

We demonstrate that the two criteria yield rather different results and that the AIC based on the Extended Restricted Likelihood (ERL) performs better when applied to some well-known multivariate survival data sets. Finally, we discuss some extensions of the method to more general models.

## References

[1] Ha ID, Lee, Y & MacKenzie G. (2007). Model selection for multi-component frailty models.Statistics in Medicine,26 4790-4807.