1

Extended Cox and Accelerated Models in Reliability, with General Censoring and Truncation

Catherine Huber[1] **and Mikhail Nikulin**[2]

[1] *Université Paris 5 and Unité INSERM 870, Paris, France*
[2] *EA 2961, Université de Bordeaux 2, Bordeaux, France*

Abstract: We review recent developments in reliability or survival analysis. We consider various models for the time to failure or survival time, by a law on $\mathbb{R}^+$ that may depend on one or more factors. Inhomogeneity is taken into account by way of frailty models. The presence of censoring and truncation of a general type, more complex than the usual simple case of right censoring, induced the most recent developments on these topics. In case of clusters of items or families of patients implying a possible dependence between multiple failure times, shared frailty models or hierarchical dependency models are considered.

Keywords and Phrases:

1.1 Cox Model and Extensions

Let the failure time X have survival function $S(x) = P(X \geq x)$, density $f = -dS/dx$, hazard $\lambda(x) = f(x)/S(x)$, cumulative hazard $\Lambda(x) = \int_0^\infty \lambda(u)du$.

1.1.1 The simple Cox model

The basic Cox model assumes that conditional on a p-dimensional covariate $Z = z$, the hazard rate verifies

$$\lambda(x|z) = \lambda_0(x)e^{(<\beta,z>)}$$

where β is an unknown p-dimensional parameter and λ_0 an unknown function of x. It is the most popular model because it leads to very easy interpretation of the impact of each component of the covariate, over all when they are constant in time. But it suffers some limitations. Let the covariates be constant

in time at the moment. First, the multiplicative dependence on the covariate z, assumed to be exponential could be replaced by a function $\varphi(z)$; the corresponding model is called a proportional hazard (PH) model (see Definition 1.4.4). The second limitation is that the model depends only on the time x elapsed between the starting event (for example, diagnosis) and the terminal event (for example, death), and not on the chronological time t; it is actually assumed to be homogeneous in chronological time. One could introduce a dependence on x and t. The third limitation is that the effect of a given covariate is constant in time. This leads to the fact that the survival functions $S(.|z)$ and $S(.|z')$ corresponding to two distinct values z and z' of Z are always ordered, for example, $S(x|z) < S(x|z')\,\forall x$, without any possibility of crossing. A fourth limitation is that if one pertinent covariate is omitted, even if it is independent of the other covariates in the model, averaging on the omitted covariate gives a new model that is no longer of Cox type, and if it is treated as such, this leads to (possibly very) biased estimates [Bretagnolle and Huber (1988)] of the regression coefficients β. Frailty models take care of this case, introducing heterogeneity in the Cox model.

1.1.2 Non homogeneity in chronological time

In order to take into account the effect of the initial time t, there are several possibilities: either add it, possibly categorized, as a $(p+1)$th component of covariate $Z = (Z_1, \ldots, Z_p)$ or have a baseline hazard which is both a function of x and t, $\lambda_0 = \lambda_0(x;t)$. The second proposal is due to Pons (2001, 2002) and Pons and Visser (2000) who studied its asymptotic properties allowing for the use of the following model:

$$\lambda(x|t,z) = \lambda_0(x;t)e^{<\beta,z(t+x)>}.$$

The usual Nelson–Aalen estimate for Λ is a kernel estimate [Pons and Visser (2000)], with kernel K a continuous, symmetric density, with support $[-1,+1]$ and $K_{h_n}(s) = (1/h_n)K(s/h_n)$ where $h_n \longrightarrow 0$ at a convenient rate

$$\widehat{\Lambda_{n,X|S}}(x;s;\beta) = \sum_i \frac{K_{h_n}(s-S_i)\delta_i 1\{X_i \le x\}}{nS^{(0)}}$$

where $S^{(0)}$ is defined as $S^{(0)} = (1/n)\sum_j K_{h_n}(s-S_j)Y_j(x)e^{<\beta,Z_j(S_j+x)>}$ $\widehat{\beta_n}$ maximizes the partial likelihood:

$$l_n(\beta) = \sum_i \delta_i[<\beta, Z_i(T_i^0)> - \ln\{nS^{(0)}(X_i;s;\beta)\}]\epsilon_n(S_i)$$

where $\epsilon_n(s) = 1\{s \in [h_n, \tau - h_n]$.

Good asymptotic properties of those estimators were proved. Also a Goodness of fit test was derived, as well as a test of H_0 : the model is Cox, homogeneous in time: $\lambda_{X|S}(x;s) \equiv \lambda_X(x)$ against H_1: the model depends also on chronological time: $\lambda_{X|S}(x;s)$.

1.1.3 Effect not constant in time

The simplest way to involve effects not constant in time of some covariates is to consider that β is actually piecewise constant. This can be viewed as a breakpoints problem with the step function $\beta(x)$ having an unknown number k of steps (known to be bounded by a fixed constant k_0) as well as an unknown localization of the jumps, depending on the admissible complexity of the model. Another way is to consider that the constant ratio β may vary as a function of an observed covariate $Z_0 \in I\!R$: $\beta = \beta(Z_0)$, such that $Z_0 \sim f_0$ for some density f_0. The corresponding model [Pons (2001)] is

$$\lambda(t|Z_0, Z) = \lambda_0(t)e^{<\beta(Z_0),Z(t)>}.$$

Observations are $(T_i, \delta_i, Z_{i0}, Z_i)$ $, i = 1, 2, \ldots, n;$ $T_i = T_i^0 \wedge C_i$ $; \delta_i = 1\{T_i^0 \leq C_i\}$. The problem is to estimate $\beta(z_0)$ on a compact subset J_{Z_0} of the support of Z_0. $\hat{\beta}_n(z_0)$ maximizes the partial likelihood

$$\begin{aligned} l_{n,z_0}(\beta) &= \sum_{i\leq n} \delta_i K_{h_n}(z_0 - Z_{i0})[<\beta, Z_i(T_i)> \\ &\quad - \ln\{\sum_{j\leq n} K_{h_n}(z_0 - Z_{j0})Y_j(T_i)e^{<\beta,Z_j(T_i)>}\}]\widehat{\Lambda_n}(t) \\ &= \sum_{i:T_i\leq t} \frac{\delta_i}{S_n^{(0)}(T_i)} \end{aligned}$$

where $S_n^{(0)}(s) = \sum_j Y_j(s)1\{Z_{j_0} \in J_{Z_0}\}e^{<\widehat{\beta_n}(Z_{j_0}),Z_j(T_i)>}$.

1.1.4 Omitted pertinent covariate: frailty models

Let the Cox model be true, but one component of the covariate Z is omitted or unobserved, say the $(p+1)^{th}$ component. $S(t|Z' = (z_1, \ldots, z_p))$ is equal to $S(t|Z = (z_1, \ldots, z_{p+1}))$ averaged on z_{p+1}. Denoting

$$\eta = e^{\beta_{p+1}z_{p+1}},$$

the corresponding model is a frailty model thus defined: η is a positive random variable, the survival of subject i, $i = 1, \cdots, n$, whose p-dimensional covariate z_i is observed and frailty η_i is not observed, but has known distribution function F_η on $I\!R^+$. The X_is are independent and their survival function S and hazard

function h obey the following frailty model, where $\beta \in R^p$ is an unknown regression parameter, $\lambda(t)$ the unknown baseline hazard and $\Lambda(t) = \int_0^\infty \lambda(u)du$ the baseline cumulative hazard:

$$\begin{aligned} h(t|z,\eta) &= \eta e^{\beta^T z}\lambda(t) \quad (1.1) \\ S(t|z,\eta) &= e^{-\eta e^{\beta^T z}\Lambda(t)} \end{aligned}$$

$$S(t|z) = \int_0^\infty e^{-xe^{\beta^T z}\Lambda(t)} dF_\eta(x) = e^{-G(e^{\beta^T z}\Lambda(t))} \quad (1.2)$$

where G is equal to $-\log$ of the Laplace transform of η:

$$G(y) = -\ln(\int_0^\infty e^{-uy} dF_\eta(u)) \quad (1.3)$$

The two most popular frailty distributions are the Gamma (Clayton-Cuzick frailty model) with mean 1 and variance c, and the Inverse Gaussian with mean 1 and variance $1/2b$. The respective functions G defined in (1.3) are equal to:

$$\begin{aligned} G(x,c) &= \tfrac{1}{c}\ln(1+cx),\ c>0 \\ G(x,b) &= \sqrt{4b(b+x)} - 2b,\ b>0. \end{aligned}$$

1.2 General Censoring and Truncation

1.2.1 Definition

Very often, the failure or survival time X is right censored and classical statistical inference is obtained under this assumption. But it may also happen rather frequently that X is both censored, in a more general way than on its right, and also truncated, so that the X_is are generally not observed. Instead, one observes two intervals (A_i, B_i), which are respectively the censoring interval, $A_i = [L_i; R_i]$, and the truncating interval $]\mathcal{L}_i; \mathcal{R}_i[$, such that $B_i \supset A_i$. This means that X_i is not observed but is known to lie inside A_i, and A_i itself is observed only conditionally on the fact that it is inside the truncating interval B_i. Otherwise, the corresponding subject is said to be "truncated," i.e., it does not appear in the sample. Finally, for the n subjects who are not truncated, the observations are (A_i, B_i, z_i), $i \in \{1, 2, \cdots, n\}$. When in model (1.2), there is no covariate and G is the identity, the non parametric maximum likelihood estimate under general censoring and truncation is due to the early work of Turnbull (1976). It was then extended to the semi-parametric Cox model by Alioum and Commenges (1994), and to the general frailty model (1.2) by Huber and Vonta (2004). The consistency of the NPML estimate of the density of X was proved [Huber, Solev, and Vonta (2006)] under regularity conditions on the laws of X and of the censoring and truncation schemes.

1.2.2 Maximum likelihood estimation for frailty models

Under the above censoring and truncation scheme, the likelihood is proportional to

$$l(S) = \prod_{i=1}^{n} l_i(S_i) = \prod_{i=1}^{n} \frac{P_{S_i}(A_i)}{P_{S_i}(B_i)} = \prod_{i=1}^{n} \frac{\left\{S_i(L_i^-) - S_i(R_i^+)\right\}}{\left\{S_i(\mathcal{L}_i^+) - S_i(\mathcal{R}_i^-)\right\}}. \tag{1.4}$$

Following Turnbull (1976), we define the "beginning" set $\tilde{L}$ and the "finishing" set $\tilde{R}$, in order to take advantage of the fact that the likelihood is maximum when the values of $S_i(x)$ are the greatest possible for $x \in \widetilde{L}$ and the smallest possible for $x \in \widetilde{R}$:

$$\begin{aligned} \widetilde{L} &= \{L_i, 1 \le i \le n\} \cup \{\mathcal{R}_i, 1 \le i \le n\} \cup \{0\} \\ \widetilde{R} &= \{R_i, 1 \le i \le n\} \cup \{\mathcal{L}_i, 1 \le i \le n\} \cup \{\infty\}. \end{aligned}$$

Let

$$Q = \{[q'_j p'_j] : q'_j \in \widetilde{L} \ , \ p'_j \in \widetilde{R} \ , \ [q'_j p'_j] \cap \widetilde{L} = \varnothing \ , \ [q'_j p'_j] \cap \tilde{R} = \varnothing\}$$

$$0 = q'_1 \le p'_1 < q'_2 \le p'_2 < \ldots < q'_v \le p'_v = \infty.$$

Then,

$$Q = \cup_{j=1}^{v} [q'_j, p'_j] = C \cup W \cup D$$

where

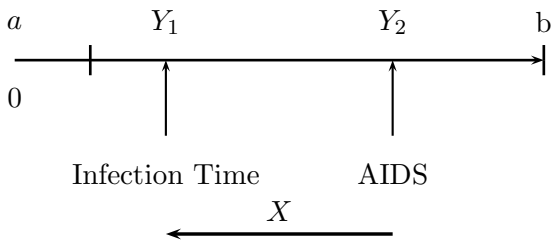
$$\begin{array}{rcll} C & = & \cup[q'_j, p'_j] & \text{covered by at least one censoring set,} \\ W & = & \cup[q'_j, p'_j] & \text{covered by at least one truncating set,} \\ & & & \text{but not covered by any censoring set,} \\ D & = & \cup[q'_j, p'_j] & \text{not covered by any truncating set.} \end{array}$$

The special case of $G \equiv Id$ and $\beta = 0$ was studied in detail in Turnbull (1976), followed by Frydman (1994) and Finkelstein (1993).

The above likelihood, for the general frailty model (1.2) as a function of the unknown β and Λ, is equal to

$$l(\Lambda, \beta | (A_i, B_i, z_i)_{i \in \{1,..,n\}}) = \prod_{i=1}^{n} \frac{\left\{e^{-G(e^{\beta^T z_i} \Lambda(L_i^-))} - e^{-G(e^{\beta^T z_i} \Lambda(R_i^+))}\right\}}{\left\{e^{-G(e^{\beta^T z_i} \Lambda(\mathcal{L}_i^+))} - e^{-G(e^{\beta^T z_i} \Lambda(\mathcal{R}_i^-))}\right\}}.$$

As in the special case where $G = Id$ and $\beta = 0$ in (1.2), the NPML estimator of Λ for the frailty model (1.1) is not increasing outside the set $C \cup D$ [Huber and Vonta (2004)]. Moreover, conditionally on the values of $\Lambda({q_j}^-)$ and $\Lambda({p_j}^+)$, $1 \le j \le m$, the likelihood does not depend on how the mass $\Lambda({p_j}^+) - \Lambda({q_j}^-)$ is distributed in the interval $[q_j, p_j]$. From this remark, follows the object of the estimation of Λ and β. The special case of $G = Id$ was studied by Alioum and Commenges.



$X = Y_2 - Y_1$: AIDS induction time.

Figure 1.1: AIDS example of right truncation

1.3 Discrete Time: Logistic Regression Models for the Retro-Hazard

A very classical example of censored and truncated survival data is the retrospective AIDS induction time for patients infected by blood transfusion [Kalbfleish (1989)]. The starting date Y_1, infection time, is reached retrospectively from Y_2, the time of onset of AIDS. $0 < Y_1 + X \leq b$ holds, which means that X is right truncated by $b - Y_1$. When, moreover, one knows that Y_1 took place *after* the first transfusion, Y_0, X may be also left censored by $Y_2 - Y_0$. We have there a censoring variable $C = Y_2 - Y_0$ and a truncating variable $T = b - Y_1$. We have there data that are left censored and right truncated. The treatment of this kind of data is the same as the treatment of right censored left truncated data, that is implied hereafter.

Assuming now that, for those censored and truncated data, time is discrete, with values $\{1, 2, \ldots, k\}$. X is the survival, C the right censoring and T the left truncating variable, they are independent and the model for X is the logistic model for the retro-hazard $h^*(t)dt = [f(t)dt]/[1 - S(t)]$:

$$\log \frac{(h^*(t|Z(t) = z)}{(1 - h^*(t|Z(t) = z))} =< \beta, z >, \qquad t \in \{1, 2, \ldots, T\}.$$

Gross and Huber (1992) obtain non parametric estimators and tests for the saturated model when all covariates are categorical, for the three laws of X, the survival, C the censoring and T the truncation, using a special partial likelihood. In Figure 1.1, observations take place in the hatched triangle, due to left truncation, and the risk set at time i is the hatched rectangle.

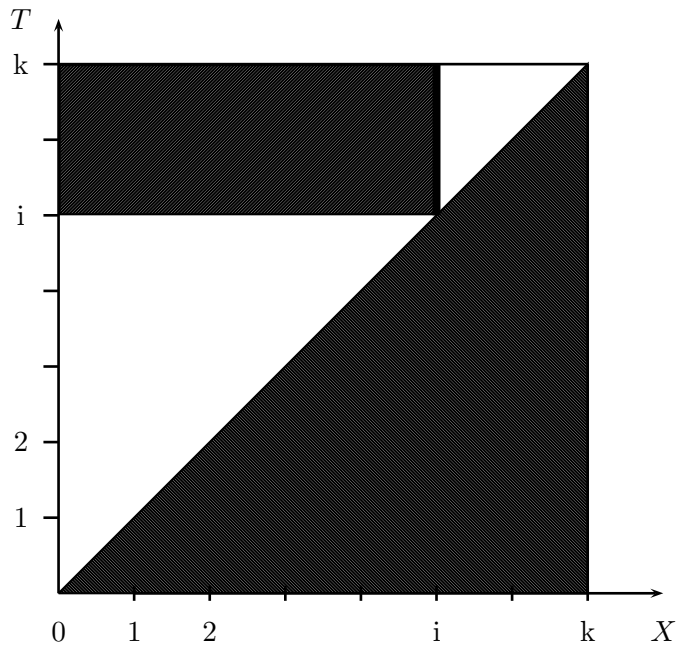


Figure 1.2: Risk zone for right censored and left truncated discrete times

1.4 Accelerated Failure Time Models (AFT)

Enforced controlled stresses are meant to reduce the time on test. It is used in particular for tires, brakes and more generally for planes and trains equipments. Hence the need for a transfer functional [Nikulin (????)] allowing an interpolation from the time to failure under enforced stress to the time to failure under regular stress: Sedyakin principle.

1.4.1 Sedyakin principle

Let $\mathcal{E}_1$ be the set of constant stresses, $\mathcal{E}_2$ the step stresses, thus defined:

$$\mathcal{E}_2 = \{Z(.) : Z(t) = Z_1 \ \mathbb{1}\{0 \leq t \leq t_0\} + Z_2 \ \mathbb{1}\{t > t_0\}; Z_1, Z_2 \in \mathcal{E}_1\}.$$

The Sedyakin principle may then be formulated as follows for step stresses:

Definition 1.4.1 (Sedyakin principle (A_S) **on** $\mathcal{E}_2$**)** Let $Z_1(.)$ and $Z_2(.)$ be two stresses. We say that $t_1 \sim t_2$ if $S(t_1|Z_1(.)) = S(t_2|Z_2(.))$. If $Z_1(.) = Z_1$ constant, $Z_2(.) = Z_2$ constant and $Z(t) = Z_1 \ \mathbb{1}\{0 \leq t \leq t_1\} + Z_2 \ \mathbb{1}\{t > t_1\}$, then Sedyakin principle (A_S) on $\mathcal{E}_2$ holds if

$$\lambda(t_1 + s|Z(.)) = \lambda(t_2 + s|Z_2).$$

Let $\mathcal{E}$ be the set of the general stresses, that are p-dimensional left continuous processes having right limits. Then, the Sedyakin principle for general stress is

Definition 1.4.2 (Generalized Sedyakin principle (A_{GS}) on $\mathcal{E}$) A model obeys generalized Sedyakin assumption (A_{GS}) if there exists a function g such that

$$\lambda(t|Z(s)\ 0 \leq s \leq t) = g(Z(t), S(t|Z(s)\,; 0 \leq s \leq t)).$$

It means that the hazard rate $\lambda(t|Z(.))$ is independent on the past conditionally on $\Lambda(t|Z(s), 0 \leq s < t))$:

$$(\lambda(t|Z(.)) \overset{\Lambda(t|Z(.))}{\perp} \mathcal{F}_{t^-}$$

or equivalently on $S(t|Z(s), 0 \leq s < t))$ sometimes called the *resource*.

1.4.2 Definition of AFT models

Loosely speaking, an accelerated model is a model based on a given survival function G and a transformation $\alpha(t)$ of time t, where α is a non decreasing function: $S(t) = G(\alpha(t))$. This acceleration $(\alpha > Id)$ or deceleration $(\alpha < Id)$ takes place through a positive function r of the stress $Z(s)$; $0 < s \leq t$:

Definition 1.4.3 (AFT model on $\mathcal{E}$) A model is AFT on $\mathcal{E}$ if there exists a survival function G and a positive function r such that:

$$S(t|Z(s), 0 \leq s \leq t) = G(\int_0^t r(Z(s))ds) \quad \forall Z \in \mathcal{E}.$$

In the simple case of a constant stress $Z \in \mathcal{E}_1 : Z = z_0$:

$$S(t|Z) = G(r(z_0)t) \quad \forall Z \in \mathcal{E}_1. \qquad (*)$$

There is a relationship between Sedyakin $(A_{GS})$and AFT models [Bagdonavicius and Nikulin (2002)]: A_{GS} and $(*)$ hold $\Longleftrightarrow \exists q > 0, r > 0$ such that

$$\lambda(t|Z(.)) = r(Z(t)) * q(S(t|Z(.)).$$

An AFT model on $\mathcal{E}_2$ is such that if Z_1 and Z_2 are constant stresses, and $Z(t) = Z_1\ \mathbb{1}\{0 \leq t \leq t_1\} + Z_2\ \mathbb{1}\{t > t_1\}$, then

$$t_2 = \frac{r(Z_1)}{r(Z_2)} \sim t_1$$

$$S(t|Z(.)) = \begin{cases} S(t|Z_1), & 0 \leq t < t_1 \\ S(t - t_1 + t_2|Z_2), & t \geq t_1 \end{cases}.$$

1.4.3 Relationships between accelerated (AFT) and proportional hazard (PH) models

The Cox model is a particular case of the more general proportional hazard (PH) models:

Definition 1.4.4 (PH model) A PH model on $\mathcal{E}$ is such that, for two positive functions r and λ_0, the hazard rate verifies:

$$\lambda(t|Z(.)) = r(Z(t))\ \lambda_0(t)\ \forall Z(.) \in \mathcal{E}.$$

Then $\Lambda(t|Z(.)) = \int_0^t r(Z(s))d\Lambda_0(s)$ and $S(t|Z(.)) = e^{-\int_0^t r(Z(s))d\Lambda_0(s)}$, where $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$, and $S_0(t) = e^{-\Lambda_0(t)}$. The simple case of a PH model on $\mathcal{E}_1$ gives $\lambda(t|Z) = r(Z)\lambda_0(t) \quad \forall Z \in \mathcal{E}_1$. The corresponding survival is then $S(t|Z) = S_0^{r(Z)}(t) = e^{-r(Z)\Lambda_0(t)}$. Let $\rho(Z_1, Z_2) = r(Z_2)/r(Z_1)$. Then $S(t|Z_2) = S(t|Z_1)^{\rho(Z_1,Z_2)}$. If PH holds on $\mathcal{E}_2$, then $\forall Z(.) \in \mathcal{E}_2$ such that for two constant stresses Z_1 and Z_2, $Z(t) = Z_1\ \mathbb{1}\{0 \le t \le t_1\} + Z_2\ \mathbb{1}\{t > t_1\}$,

$$\lambda(t|Z(.)) = \begin{cases} \lambda(t|Z_1) = r(Z_1)\lambda_0(t), & 0 \le t \le t_1 \\ \lambda(t|Z_2) = r(Z_2)\lambda_0(t), & t > t_1 \end{cases}$$

and

$$S(t|Z(.)) = \begin{cases} S(t|Z_1), & 0 \le t \le t_1 \\ S(t|Z_2)\frac{S(t_1|Z_1)}{S(t_1|Z_2}, & t > t_1 \end{cases}.$$

1.4.4 Relationships between Sedyakin and PH: MPH models

Bagdonavicius and Nikulin (2002) define a proportional hazard model that obeys the Sedyakin principle:

Definition 1.4.5 (Modified PH model: MPH) A model is Sedyakin (A_{GS}) on $\mathcal{E}$ and PH on $\mathcal{E}_1$ and called MPH if and only if, for two functions r and λ_0,

$$\lambda(t|Z(.)) = r(Z(t))\lambda_0\left(\Lambda_0^{-1}\left(\frac{\Lambda(t|Z(.))}{r(Z(t))}\right)\right).$$

If MPH holds on $\mathcal{E}_2$, then $\forall Z(.) \in \mathcal{E}_2$ such that for two constant stresses Z_1 and Z_2:

$$\begin{aligned} Z(t) &= Z_1\ \mathbb{1}\{0 \le t \le t_1\} + Z_2\ \mathbb{1}\{t > t_1\} \\ t_2 &= S^{-1}((S(t_1, Z_1))^{\rho(Z_2,Z_1)}) \end{aligned}$$

then

$$S(t|Z(.)) = \begin{cases} S(t|Z_1), & 0 \le t < t_1 \\ S(t - t_1 + t_2|Z_2), & t \ge t_1 \end{cases}.$$

1.4.5 Generalized PH models (GPH) on $\mathcal{E}$

Two distinct generalized PH models are defined, GPH1 and GPH2:

Definition 1.4.6 (GPH1) A model is GPH1if and only if, for two positive functions r and λ_0, the hazard λ verifies

$$\lambda(t|Z(.)) = r(Z(t)) * q\{\Lambda(t, Z(.))\} * \lambda_0(t)$$

When $q \equiv 1$ this is simply a PH model, while $\lambda_0(t) \equiv \lambda_0$ constant gives the AFT model.

Definition 1.4.7 (GPH2) A model is GPH2 if and only if, for two positive functions u and λ_0,

$$\lambda(t|Z(.)) = u(Z(t), \Lambda(t|Z(.)) * \lambda_0(t).$$

$\lambda_0(t) \equiv \lambda_0$ constant gives a GS model on $\mathcal{E}$, while $u(Z, s) = r(Z)q(s)$ gives a GPH1 model. Model GPH1 holds on $\mathcal{E}$ if and only if there exist two survival functions G and S_0 such that

$$S(t|Z(.)) = G\left\{\int_0^t r(Z(s))dH(S_0(s))\right\}$$

where $H = G^{-1}$. Function f_G defined as

$$f_G(t|Z(.)) = H(S(t|Z(.))$$

is called the *transfer functional*. It is the G-resource used until time t under stress Z. It is actually a transfer of quantiles.

1.4.6 General models

There are many relationships between those models. One can construct a general model that contains most of the models defined above.

1. Accelerated model (AFT)

$$\lambda(t|Z) = r(Z)q\{S(t|Z)\}$$

2. Generalized proportional models of type 1 (GPH1)

$$\lambda(t|Z) = r(Z)q\{\Lambda(t)\}$$

include the following sub-models:

$$\begin{array}{lcll} q(v) & = & 1 & (PH) \\ q(v) & = & (1+v)^{\gamma+1} & \\ q(v) & = & e^{\gamma v} & \\ q(v) & = & \frac{1}{(1+\gamma v)} & \end{array}$$

3. Generalized proportional models of type 2 (GPH2)

$$\lambda(t|Z) = u\{Z, \Lambda(t|Z)\}\lambda_0(t)$$

whose sub-models correspond to various choices of function u. Statistical inference on those various models may be found in Bagdonavicius and Nikulin (2002).

1.4.7 Modeling and homogeneity problem

General models, considered here, are very useful not only for construction of goodness-of-fit tests for the PH model but also they give the possibility to construct goodness-of-fit tests for data homogeneity hypothesis. Following Bagdonavicius and Nikulin (2005) we give here three models each of them including the PH model.

Generalized proportional hazards (GPH) model on $\mathcal{E}_1$:

$$\lambda(t,|Z) = e^{\beta^T Z}(1 + \gamma e^{\beta^T Z}\Lambda_0(t))^{\frac{1}{\gamma}-1} * \lambda_0(t).$$

This model has the following properties on $\mathcal{E}_1$: the ratios of the hazard rates increase, decrease or are constant, the hazard rates and the survival function do not intersect in the interval $(0, \infty)$. *Simple cross-effects (SCE) model* $\mathcal{E}_1$ [Bagdonavičius and Nikulin (2005)]

$$\lambda(t,|Z) = e^{\beta^T Z}\{1 + e^{(\beta+\gamma)^T}\Lambda_0(t)\}^{e^{-\gamma^T Z}-1} * \lambda_0(t).$$

The SCE model has the following properties on $\mathcal{E}_1$: the ratios of the hazard rates increase, decrease or are constant, the hazard rates and the survival function do not intersect or intersect once in the interval $(0, \infty)$. *Multiple cross-effects (MCE) model* $\mathcal{E}_1$:

$$\lambda(t,|Z) = e^{\beta^T Z}\left(1 + \gamma^T Z\Lambda_0(t) + \delta^T Z\Lambda_0^2(t)\right)\lambda_0(t).$$

The MCE model has the next properties on $\mathcal{E}_1$: the ratios of the hazard rates increase, decrease or are constant, the hazard rates and the survival function do not intersect, intersect once or twice in the interval $(0, \infty)$.

The parameter γ is one-dimensional for the GPH model and m-dimensional for the SCE model, the parameter δ is m-dimensional.

The PH model is a particular case with $\gamma = 1$ (GPH), $\gamma = 0$ (SCE), $\delta = \gamma = 0$ (MCE).

The homogeneity (no lifetime regression) takes place if $\gamma = 1$, $\beta = 0$ (GPH), $\gamma = 0$, $\beta = 0$ (SCE), $\beta = \delta = \gamma = 0$ (MCE).

At the end let us consider the so called Hsieh models (2001), which is also a SCE model.

According to the idea of Hsieh one possible way to obtain a cross-effect of hazard rates is to take a power function of Λ_0:

$$\Lambda_x(t) = r(x_1)\Lambda_0^{\rho(x_2)}(t), \quad \lambda_x(t) = r(x_1)\rho(x_2)\Lambda_0^{\rho(x_2)-1}(t)\lambda_0(t),$$

where $x = (x_1, x_2), x_1, x_2 \in E_1, r(\cdot), \rho(\cdot) : E \to R_+^1$. Using natural parametrization $r(x_1) = e^{\beta^T x_1}$ and $\rho(x_2) = e^{\gamma^T x_2}$ we have the model

$$\Lambda_x(t) = e^{\beta^T x_1}\Lambda_0^{e^{\gamma^T x_2}}(t).$$

In particular case $x_1 = x_2 = x$ the obtained model is

$$\lambda_x(t) = e^{(\beta+\gamma)^T x}\Lambda_0^{e^{\gamma^T x}-1}(t)\lambda_0(t).$$

For any two covariates x, y the ratio $\lambda_x(t)/\lambda_y(t)$ is increasing from 0 to ∞ or decreasing from ∞ to 0. So we have a cross-effect of the hazard rates. See, also Wu (2004).

1.5 Correlated Survivals

1.5.1 Introduction

Let us first present several examples of data having the structure of correlated survival data. In diabetic retinopathy, the cluster is constituted by each diabetic patient. The survival time is the time to blindness onset for each eye separately. Two types of covariates may have an impact on the time to onset: the treatment, called a structural covariate, cluster covariates like sex and age, and individual covariates like past history of each eye. The time to onset is censored by death prior to blindness. In animal experiments on litters of rats, each litter is a cluster, and the treatment is a supposed carcinogenic product injected regularly to each rat. The survival time is the time to onset of a tumor. Again the structural covariate is the treatment, the individual covariates are sex, age, weight. The censoring takes place when death occurs before the onset of a tumor. In genetic epidemiology, the cluster is a pair of twins or a family. The survival time is the age at onset of a specific chronic disease. The structural covariates are the position inside the family (father, mother, male sib,..) and individual covariates are sex, and so on. Time is again censored by death or lost to follow up. The following picture illustrates the data structure.

cluster 1 ... cluster k .....cluster K
n_1..... n_k ..... n_K

Classical proposals to take into account the correlation induced by the clusters are frailty or copula models. There are two kinds of frailty well distinguished by Parner (1998). First, the *shared frailty*, more appropriate for taking into account inhomogeneity than dependence, as it gives the possibility of estimating the frailty distribution parameter when only one of two twins is observed. And the *shared and correlated frailty*. Gross and Huber (2002) proposed a logistic like family model in discrete time, related to hierarchical log-linear models which is detailed in the next subsection.

1.5.2 Model in discrete time: hierarchical dependencies

X_{ki} is the survival time of subject i in cluster k, and C_{ki} is the associated right censoring time. Actually, what is observed is:

$$\begin{aligned} T_{ki} &= X_{ki} \wedge C_{ki} \quad \text{observed duration.} \\ D_{ki} &= I\{X_{ki} \leq C_{ki}\} \text{ death indicator.} \end{aligned}$$

Globally data are summarized by the pair of variables (T, D) or else the two-dimensional process (R, Y), where R is the couple of the "at risk" and event processes:

$$\begin{aligned} T &= \{T_{ki};\ 1 \leq k \leq K\ ;\ 1 \leq i \leq n_k\} \\ D &= \{D_{ki};\ 1 \leq k \leq K\ ;\ 1 \leq i \leq n_k\} \end{aligned}$$

$$\begin{aligned} R_{ki}(t) &= \begin{cases} 1 & if \qquad T_{ki} \geq t \\ 0 & otherwise \end{cases} \\ Y_{ki}(t) &= \begin{cases} 1 & if \qquad D_{ki}T_{ki} = t \\ 0 & otherwise \end{cases} . \end{aligned}$$

In case of discrete time, if N is the maximum size of the clusters, data are summarized through two arrays of 0 and 1, of dimension $T \times N \times K$: $R_{T\times N\times K}$, the at risk array and the event array $Y_{T\times N\times K}$.

1.5.3 Definition of the models

The model parameters are $p_{r,y}(t)$ such that

$$P(Y = y|R = r; t) = \frac{1}{c(r,t)} \exp\left\{ \sum_{\substack{0<r'\leq r \\ 0\leq y'\leq r'\wedge y}} p_{r',y'}(t) \right\}$$

and the normalization constant is c(r,t):

$$c(r,t) = 1 + \sum_{s'\leq r} \exp\left\{ \sum_{\substack{0<r'\leq r \\ 0<y''\leq r'\wedge y'}} p_{r',y''}(t) \right\}.$$

Each model is characterized by the set $\mathcal{R}$ of those parameters $p_{r,y}$ that are equal to 0 and is thus denoted $\mathcal{H}(\mathcal{R})$; the saturated model is the one for which this set is the empty set $\emptyset$. Specially interesting are the so-called hierarchical models defined below.

Definition 1.5.1 (Hierarchical models $\mathcal{H}(\mathcal{R})$) A model is said hierarchical if all $p_{r,y}$ such that $r \notin \mathcal{R}$ are equal to 0, where $\mathcal{R}$ is a family of subsets of $\{1, 2, \ldots, N\}$, such that for any R in $\mathcal{R}$ and $R' \subset R$, $R' \in \mathcal{R}$.

Definition 1.5.2 (Model of order k, $\mathcal{H}_k$) If all $p_{r,y}$ such that $\sum_i r_i > k$ are equal to 0, the corresponding model is called model of order k and called $\mathcal{H}_k$, as all interactions up to order k are included, while interactions of order greater than k are excluded.

Models $\mathcal{H}_k$ are a special case of hierarchical models. More generally, a model may be defined by the pair $(\mathcal{R}, \mathcal{Y})$ such that $p_{r,y} = 0$ except if $r \in \mathcal{R}$ and $y \in \mathcal{Y}$, $y \subset \mathcal{Y}$.

1.5.4 Regression model

Let us now include covariates in the models:

The $p_{r,y}$ are modeled linearly in terms of time t and individual profiles, and the partial likelihood is a function of the following counts of clusters at time t,

$$\begin{aligned} N(r,y,t) &= \text{count of clusters s.t.} \begin{cases} \text{risk set} = r \\ \text{jump set} = y \end{cases} \\ N(r,t) &= \text{count of clusters s.t. risk set} = r. \end{aligned}$$

1.5.5 Estimation

Theorem 1.5.1 (Sufficient statistics) *Under the general model of dependence $(\mathcal{R}, \mathcal{Y})$ and some regularity conditions fulfilled, the sufficient statistics for the parameters of the model are the counts $N(r,t)$ and $N(r,y,t)$ for $t \in \{1, 2, \ldots, T\}$ and $(r,y) \in \mathcal{R} \otimes \mathcal{Y}$.*

Under right censoring, the same counts are the only statistics involved in the partial likelihood. One can prove consistency and asymptotic normality: $p^* =$ true set of p parameters $(\mathcal{R}^*, \mathcal{Y}^*) \subset (\mathcal{R}, \mathcal{Y}) \equiv (r^*, y^*)$ combinations, for which $P^*(R = r^*)$ and $P^*(Y = y^* | R = r^*, t)$ are strictly positive.

$\Sigma =$ matrix of the second derivatives of the log-likelihood with respect to the parameters p_{r^*,y^*}, whose general entry, for $p_{r_0^*,y_0^*}(t), p_{r_1^*,y_1^*}(t)$, is, dropping *

$$\begin{aligned} &\Sigma_{p_{r_0,y_0}(t), p_{r_1,y_1}(t)} \\ &= \sum_{\{r:\, r \geq r_0 \vee r_1\}} N(r,t)\{P\{Y(t) \geq y_0 \vee y_1) | R(i) = r\} \\ &\quad - P\{Y(t) \geq y_0 | R(t) = r\} P\{Y(t) \geq y_1 | R(t) = r\}\}. \end{aligned}$$

As the number of clusters, K, tends to infinity,

$$\frac{N(r,t)}{K} \xrightarrow{a.s.} P^*(r,t),$$

the true probability that $R(t) = r$. Similarly,

$$\frac{N(r,y,t)}{K} \xrightarrow{a.s.} P^*(r,y,t),$$

the true joint probability that $R(t) = r$ and $Y(t) = y$, for $t \in \{1, 2, \ldots, T\}$. Consequently:

$$\frac{\Sigma}{K} \xrightarrow{a.s.} \Sigma^*,$$

with typical entry:

$$\begin{aligned}
&\Sigma^*_{p_{r_0,y_0}(t),p_{r_1,y_1}(t)} \\
&\quad = \frac{1}{K} \sum_{\{r:\, r \geq r_0 \vee r_1\}} P^*(r,t)\{P^*\{Y(i) \geq y_0 \vee y_1)|R(t) = r\} \\
&\qquad -P^*\{Y(t) \geq y_0|R(t) = r\}P^*\{Y(t) \geq y_1|R(t) = r\}\} \,.
\end{aligned}$$

Theorem 1.5.2 (Consistency and as. normality) *If*

1. *for all (r', y') included in the model there exists a pair (r, y), also included in the model and such that $r \supseteq r'$ and $s \supseteq s'$, and $P^*(R = r)$ and $P^*(Y = y|R = r, t)$ are strictly positive,*

2. *Σ is nonsingular in a neighborhood of the true value p^* of the parameters,*

as K tends to infinity, the partial likelihood estimates $\widehat{p_K}$ of the parameters p is consistent and asymptotically normal:

$$\sqrt{K}(\widehat{p_K} - p_0) \xrightarrow{\mathcal{L}} N(0, \Sigma^{*^{-1}}).$$

References

1. Alioum, A., and Commenges, D. (1994). A proportional hazards model for arbitrarily censored and truncated data, *Biometrics*, **52**, 512–524.

2. Andersen, P. K., and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study, *Annals of Statistics*, **10**, 1100–1120.

3. Andersen, P. K., Borgan, O., Gill R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer Series in Statistics.

4. Bagdonavicius, V., and Nikulin, M. (1997a). Some rank tests for multivariate censored data, In *Advances in the Theory and Practice of Statistics, Volume in Honor of Samuel Kotz*, pp. 193–208, John Wiley & Sons, New York.

5. Bagdonavicius, V., and Nikulin, M. (1997b). Transfer functionals and semiparametric regression models, *Biometrika*, **84**, 365–378.

6. Bagdonavicius, V., and Nikulin, M. (1998). *Additive and multiplicative semiparametric models in accelerated life testing and survival analysis*, Queen's papers in Pure and Applied Mathematics, Queen's University, Kingston, Ontario, Canada.

7. Bagdonavicius, V., and Nikulin, M. (1999a). Generalized proportional hazards model based on modified partial likelihood, *LIDA*, **5**, 323–344.

8. Bagdonavicius, V., and Nikulin, M. (1999b). On goodness of fit for the linear transformation and frailty models, *Statistics and Probability Letters*, **6**.

9. Bagdonavicius, V., and Nikulin, M. (1999c). On nonparametric estimation in accelerated experiments with step stresses, *Statistics*, **4**.

10. Bagdonavicius, V., and Nikulin, M. (2002). Accelerated life models. Modeling and statistical analysis, In *Monographs on Statistics and Applied Probability 94*, Chapman and Hall, 344 p.

11. Bagdonavicius, V., and Nikulin, M. (2005). Statistical analysis of survival and reliability data with multiple crossings of survival functions, *Comptes Rendus de l'Academie des Sciences de Paris, Ser. I*, **340**, 377–382.

12. Brémaud, P., and Jacod, J. (1977). Processus ponctuels et martingales: résultats récents sur la modélisation et le filtrage, *Advances in Applied Probability*, **9**, 362–416.

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

14. Cai, J., and Prentice, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data, *Biometrika*, **82**, 151–164.

15. Commenges, D., and Gauzère, (1999). Non-parametric estimation in the model healthy, ill, dead, *LIDA*, **5**, 4.

16. Chang, I. S., and Hsiung, C.A. (1994). Information and asymptotic efficiency in some generalized proportional hazards models for counting processes, *Annals of Statistics*, **22**, 1275–1298.

17. Cheng, S. C., Wei, L. J., and Ying Z. (1995). Analysis of transformation models with censored data, *Biometrika*, **82**, 835–846.

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

19. Clegg, L. X., Cai, J., and Sen, P. K. (1999). A marginal mixed baseline hazards model for multivariate failure time data, *Biometrics*, **55**,3, 805–812.

20. Crowder, M. (2001). *Classical Competing Risks*, Chapman and Hall.

21. Dabrowska, D. M., and Doksum, K. A. (1988). Partial likelihood in transformation model with censored data, *Scandinavian Journal of Statistics*, **15**, 1–23.

22. De Gruttola, V., and Lagakos, S. W. (1989). Analysis of doubly-censored data with application to AIDS, *Biometrics*, **45**, 1–11.

23. Efron, B., and Petrozian, V. (1994). Survival analysis of the gamma ray burst data asymptotic efficiency in some generalized proportional hazards models for counting processes, *Journal of the American Statistical Association*, **88**, 452–462.

24. 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.

25. Finkelstein, D. M., Moore, D. F., and Shoenfeld, D. A. (1993). A proportional hazards model for truncated AIDS data, *Biometrics*, **49**, 731–740.

26. Frydman, H. (1994). A note on nonparametric estimation of the distribution function from interval-censored and truncated observations, *Journal of the Royal Statistical Society, Series B*, **56**, 71–74.

27. Genest, C., and MacKay, R. J. (1986). Archimedian copulas and bivariate families with conbtinuous marginals, *Canadian Journal of Statistics*, **14**, 145–159.

28. Genest, C., and Rivest, L. (1993). Statistical inference procedure for bivariate archimedian copulas, *Journal of the American Statistical Association*, **88**, 1034–1043.

29. Greenwood, P. E., and Wefelmeyer, W. (1993). On optimal estimating functions for partially specified counting process models, In *Estimating Functions* (Ed., V. P. Godambe), pp. 147–160, Oxford University Press.

30. Groenboom, P., and Wellner, J. (1993). *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhaüser.

31. Gross, S., and Huber, C. (1992). Regression models for truncated survival data, *Scandinavian Journal of Statistics*, **19**, 193–213.

32. Gross, S., and Huber, C. (2002). A new family of multivariate distributions for survival data, In *Goodness of Fit Tests and Model Validity*, pp. 255–266, Birkhaüser.

33. Gu, M. G., and Zhang C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data, *Annals of Statistics*, **21**, 611–624.

34. Hillali, Y. (1999). Estimation problems in arcimedian copula models, *Thesis*, INRIA, France.

35. Hougaard, P. (1987). Modeling multivariate survival, *Scandinavian Journal of Statistics*, **14**, 291–304.

36. Hsieh, F. (2001). On heteroscedastic hazards regression models: theory and application, *Journal of the Royal Statistical Society, Series B*, **63**, 63–79.

37. Huber-Carol, C., and Vonta, F. (2004). Frailty models for arbitrarily censored and truncated data, *Lifetime Data Analysis*, **10**, 369–388.

38. Huber-Carol, C., Solev, V., and Vonta, F. (2006). Frailty models for censored and truncated data, In *BIOSTAT2006, Proceedings Of the International Conference on Statistical Models in Biomedical and Technical Systems*, pp. 13–20.

39. Joly, P., Commenges, D., and Letenneur, L. (1998). A penalized likelihood approach for arbitrarily censored and truncated data: application to age-specific incidence dementia, *Biometrics*, **54**, 185–194.

40. Joly, P., and Commenges, D. (1999). A penalized likelihood approach for a progressive three-state model with censored and truncated data: application to AIDS, *Biometrics*, **55**, 887–890.

41. Kalbfleisch, J. D., and Lawless, J. F. (1989). Inference based on retrospective ascertainment: An analysis of data on transfusion-associated AIDS, *Journal of the American Statistical Association*, **84**, 360–372.

42. Lawless, J. (1982). *Statistical Models and Methods for Survival Data Analysis*, John Wiley & Sons, New York.

43. Leconte, E., Moreau, T., and Lellouch, J. (1994). The two-sample problem with multivariate censored data: a new rank test family, *Communications in Statistics—Simulation and Computation*, **23**, 1061–1076.

44. Mahé, C. (1998). Stastical analysis of multivariate failure times data, *Thèse*, Université de Paris VI.

45. Marshall, A. W., and Olkin I. (1967). A multivariate exponential distribution, *Journal of the American Statistical Association*, **62**, 30–44.

46. Meeker, W. Q., and Escobar L. (1998), *Statistical Methods for Reliability Data*, John Wiley & Sons, New York.

47. Miller, R. G., Jr. (1981). *Survival Analysis*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics.

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

49. Murphy, S. A., Rossini, A. J., and Van der Vaart, A. W. (1997). Maximum likelihood estimation in the proportional odds model, *Journal of the American Statistical Association*, **92**, 968–976.

50. Mykland, P. A., and Ren, J-J. (1996). Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data, *Annals of Statistics*, **24**, 1740–1764.

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

52. Nikulin, M., and Solev, V. (1999). Chi-squared goodness-of-fit test for doubly censored data, applied in survival analysis and reliability, In *Statistical and Probabilistic Models in Reliability* (Eds., I. Ionescu and N. Limnios), pp. 101–112, Birkhaüser, Boston.

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

54. Oakes, D., and Cai, L. (1993). On semi-parametric inference for modulated renewal processes, *Biometrika*,**81**, 83–90.

55. Parner, E. (1998). Asymptotic theory for the correlated gamma-frailty model, *Annals of Statistics*, **26**, 183–214.

56. Pons, O., and Visser, M. (2000). A non-stationary Cox model, *Scandinanvia Journal of Statistics*, **24**, 619–639.

57. Pons, O. (2000). Nonparametric estimation in a varying-coefficient Cox model, *Mathematical Methods in Statistics*, **9**, 376–398.

58. Pons, O. (2002). Inference in extensions of the Cox model for heterogeneous populations, In *Goodness-of-fit tests and validity of models* (Eds., C. Huber-Carol, N. Balakrishnan, M. S. Nikulin, and M. Mesbah), pp. 211–225, Birkhaüser, Boston.

59. Pons, O. (2002). Estimation in a Cox regression model with a change-point at an unknown time, *Statistics*, **36**, 101–124.

60. Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.

61. Schaffer, J. L. (1997). *Analysis of incomplete multivariate data*, Chapman and Hall, London.

62. Schweizer, B., and Sklar, A. (1983). *Probabilistic Metric Spaces*, North-Holland, Amsterdam.

63. Slud, E. V., and Vonta, F. (2004). Consistency of the NPML estimator in the right-censored transformation model, *Scandinavian Journal of Statistics*, **31**, 21–41.

64. Tsiatis, A. A. (1981). A large sample study of Cox's regression model, *Annals of Statistics*, **9**, 93–108.

65. Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data, *Journal of the American Statistical Association*, **69**, 290–295.

66. Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data, *Journal of the Royal Statistical Society*, **38**, 290–295.

67. Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality, *Demography*, **16**, 439–47.

68. Viertl, R. (1988). *Statistical Methods in Accelerated Life Testing*, Vandenhoeck, and Ruprecht in Gottingen.

69. Wei, L.-J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, *Journal of the American Statistical Association*, **84**, 1065–1073.

70. Wu, H-D.I. (2004). Effect of ignoring heterogeneity in hazards regression, In *Parametric and Semipametric Models with Applicatiions to Reliability, Survival Analysis, and Quality of Life*, pp. 239–252, Birkhaüser, Boston.

71. Zhao, L. P., and Prentice, R. L. (1990). Correlated binary regression using the quadratic exponential model, *Biometrika*, **77**, 642–648.