# Frailty Models For Arbitrarily Censored And Truncated Data

Catherine Huber *, and Filia Vonta †

May 5, 2004

**Abstract**

In this paper we propose a frailty model for statistical inference in the case where we are faced with arbitrarily censored and truncated data. Our results extend those of Alioum and Commenges (1996) who developed a method of fitting a proportional hazards model to data of this kind. We discuss the identifiability of the regression coefficients involved in the model which are the parameters of interest, as well as the identifiability of the baseline cumulative hazard function of the model which plays the role of the infinite dimensional nuisance parameter. We illustrate our method with the use of simulated data as well as with a set of real data on transfusion-related AIDS.

## 1 Introduction

A common feature of many failure time data in epidemiological studies is that they are simultaneously truncated and interval-censored. For instance, right-truncated data occur in registers. An acquired immune deficiency syndrome (AIDS) register only contains AIDS cases which have already been reported, which generates right-truncated samples of induction times. As for the interval-censoring it comes usually from grouped data or from the fact that patients are examined at certain dates and the event of interest is only known to have occured between two specific checking times, one of which may be infinite in case of right-censoring, when at the end of the study the event has not yet occured.

The most widely used model in survival analysis is the Cox proportional hazards model (Cox 1972). Although the cases of right-censored and/or left-truncated data can be handled through the standard method of estimation in the Cox model, namely, the partial likelihood, the cases for example of interval-censored or right-truncated data should be treated differently. Turnbull (1976) and Frydman (1994) dealt with the nonparametric estimation of the distribution function $F$ when the data are interval-censored

*MAP 5, FRE CNRS 2428, UFR Biomédicale, Université René Descartes, et U 472 INSERM, France, e-mail: catherine.huber@univ-paris5.fr.

†Department of Mathematics and Statistics, University of Cyprus P.O. Box 20537, CY-1678, Nicosia, Cyprus. e-mail: vonta@ucy.ac.cy.

and truncated. Discrete time regression models for right-truncated data have been developed among others by Gross and Huber-Carol (1992). Finkelstein (1986) fitted a proportional hazards model to interval-censored data and Finkelstein *et al.* (1993) to right-truncated data, applying their results in estimation and hypothesis testing on real data concerning AIDS patients. Huang (1994) and Huang and Wellner (1995) examined the theoretical aspects related to the NPMLE of the regression coefficient and the baseline distribution, in the case of the Cox model as well as in a class of semiparametric models, with interval-censoring. Alioum and Commenges (1996) extended the existing results by fitting a proportional hazards model to arbitrarily censored and truncated data and concentrated on hypothesis testing. Pan and Chappell (2002) concentrated on the estimation of the parameters involved in the Cox model with left-truncated and interval-censored data. They showed that the NPMLE can seriously underestimate the baseline survival but it works well in estimating the regression coefficient. In this paper we introduce frailty models for the case of arbitrarily censored and truncated data and focus on estimation of the parameters of interest as well as the nuisance parameter of our model.

The need to apply frailty models to analyze survival data arises when the assumption of a homogeneous population seems questionable. In order to model unobserved heterogeneity in the population one introduces a random effect into the model, called frailty, defined to act multiplicatively on the hazard rate $h(t|z)$ of an individual with covariate vector $z$. A frailty model therefore arises naturally from a Cox model with unobserved covariates which materialize the frailty parameter. A frailty parameter is also introduced to model dependence between survival times if the standard assumption of independence seems unrealistic. The concept of frailty was introduced by Vaupel *et al.* (1979) who studied the model with Gamma distributed frailties. There are many frailty distributions one could consider with the choice of Gamma being the most popular due to its mathematical convenience. This particular model is well known as the Clayton-Cuzick model (Clayton-Cuzick, 1985 and 1986). Vaupel and Yashin (1983) examined other frailty distributions such as the uniform, the Weibull and the log-normal distributions. Other choices of frailty distributions include (see for example Hougaard 1984 and 1986) the Inverse Gaussian and the Positive Stable distributions. Sahu and Dey (2003) developed a class of log-skew-t distributions for the frailty parameter. This class includes the log-normal distribution along with many other heavy tailed distributions such as the log-t and the log-Cauchy.

In section 2 we formulate the appropriate likelihood for the case of arbitrarily censored and truncated data following the notation of Turnbull (1976) and Alioum and Commenges (1996). In section 3 we define a general class of semiparametric frailty or transformation models that we will be using throughout the paper and rewrite the likelihood for this class of models. We also reorganize the log-likelihood, as has been done by previous authors, to produce a form that will be convenient for maximizaton with respect to parameters. In section 4, we discuss the identifiability of $\beta \in R^p$, the parameter of interest, as well as the identifiability of the baseline cumulative hazard function $\Lambda$, namely, the nuisance parameter. In section 5 we illustrate the performance of our model with simulated data and a set of real data on

transfusion-related AIDS. In section 6 we comment on our method and summarize all our important conclusions.

## 2 Formulation

We present here the general framework of the case of arbitrarily censored and truncated data for independent and identically distributed positive random variables following the formulation of Turnbull (1976), Frydman (1994) and especially Alioum and Commenges (1996). Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed positive random variables with survival function $S(x)$. For every random variable $X_i$ we have a pair of observations $(A_i, B_i)$ where $A_i$ is a set called the censoring set and $B_i$ a set called the truncating set. The random variable $X_i$ belongs to the sample only if $X_i$ falls into the set $B_i$. Also, $X_i$ is being censored by the set $A_i$ in the sense that the only thing that we know about $X_i$ is that it belongs to the set $A_i$ where $A_i \subseteq B_i$. The sets $A_i$ belong to a partition $\mathcal{P}_i$ of $[0, \infty)$ and we assume that $B_i$ and $\mathcal{P}_i$ are independent of $X_i$ and of the parameters of interest. We assume that the censoring sets $A_i,\ i = 1, \ldots, n$ can be expressed as a finite union of disjoint closed intervals, that is,

$$A_i = \cup_{j=1}^{k_i} [L_{ij}, R_{ij}]$$

where $0 \leq L_{i1} \leq R_{i1} < L_{i2} \leq R_{i2} < \ldots < L_{ik_i} \leq R_{ik_i} \leq \infty$ for $i = 1, \ldots, n$, $R_{i1} > 0$, $L_{ik_i} < \infty$. Moreover, we assume that the truncating sets $B_i$ can be expressed as a finite union of open intervals

$$B_i = \cup_{j=1}^{n_i} (\mathcal{L}_{ij}, \mathcal{R}_{ij})$$

where $0 \leq \mathcal{L}_{i1} < \mathcal{R}_{i1} < \mathcal{L}_{i2} < \mathcal{R}_{i2} < \ldots < \mathcal{L}_{in_i} < \mathcal{R}_{in_i} \leq \infty$ for $i = 1, \ldots, n$.

The likelihood of the $n$ pairs of observations $(A_i, B_i), i = 1, 2, \ldots, n$ is proportional to

$$l(S) = \prod_{i=1}^{n} l_i(S) = \prod_{i=1}^{n} \frac{P_S(A_i)}{P_S(B_i)} = \prod_{i=1}^{n} \frac{\sum_{j=1}^{k_i} \left\{ S(L_{ij}^-) - S(R_{ij}^+) \right\}}{\sum_{j=1}^{n_i} \left\{ S(\mathcal{L}_{ij}^+) - S(\mathcal{R}_{ij}^-) \right\}} \tag{1}$$

We are interested in defining a nonparametric maximum likelihood estimator (NPMLE) of the survival function $S$, which decreases only in a finite number of disjoint intervals. Let us define now the sets

$$\tilde{L} = \{L_{ij},\ 1 \leq j \leq k_i, 1 \leq i \leq n\} \cup \{\mathcal{R}_{ij},\ 1 \leq j \leq n_i, 1 \leq i \leq n\} \cup \{0\}$$

and

$$\tilde{R} = \{R_{ij},\ 1 \leq j \leq k_i, 1 \leq i \leq n\} \cup \{\mathcal{L}_{ij},\ 1 \leq j \leq n_i, 1 \leq i \leq n\} \cup \{\infty\}.$$

Notice that the above likelihood is maximized when the values of $S(x)$ are as large as possible for $x \in \tilde{L}$ and as small as possible for $x \in \tilde{R}$. A set $Q$ is defined uniquely as the

union of disjoint closed intervals whose left endpoints lie in the set $\tilde{L}$ and right endpoints in the set $\tilde{R}$ respectively, and which contain no other members of $\tilde{L}$ or $\tilde{R}$. Thus,

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

where $0 = q'_1 \leq p'_1 < q'_2 \leq p'_2 < \ldots < q'_v \leq p'_v = \infty$. Subsequently, we denote by $C$ the union of intervals $[q'_j, p'_j]$ covered by at least one censoring set, $W$ the union of intervals $[q'_j, p'_j]$ covered by at least one truncating set but not covered by any censoring set and $D = \overline{(\cup B_i)}$ the union of intervals $[q'_j, p'_j]$ not covered by any truncating set. $D$ is actually included in the union of intervals $[q'_j, p'_j]$. That can be proved as follows. Let $r$ be a point not covered by any truncating set and neither being a left nor a right endpoint of a truncating set. Then there exists $l$ such that $r \in [q'_l, p'_l]$ as

$$\begin{aligned} \mathcal{R}_{i_1 j_1} &= max_{i,j}\{\mathcal{R}_{ij} : \mathcal{L}_{ij} < r\} < r \\ \mathcal{L}_{i_2 j_2} &= min_{i,j}\{\mathcal{L}_{ij} : \mathcal{R}_{ij} > r\} > r \end{aligned}$$

so that $r \in [q'_l, p'_l] \equiv [\mathcal{R}_{i_1 j_1}, \mathcal{L}_{i_2 j_2}]$.

Obviously, the set $Q$ can be written as $Q = C \cup W \cup D$. The next two Lemmas that appear in Turnbull (1976) and Alioum and Commenges (1996) are essential for the maximization of (1) with respect to $S$ and become apparent upon examination of (1).

**Lemma 1** *Any survival distribution function which decreases outside the set $C \cup D$ cannot be the NPMLE of $S$.*

A first comment is therefore that a NPMLE of $S$ lies among the functions that are constant outside the set $C \cup D$. Moreover notice that from the data we can only estimate the conditional survival function $S_{\overline{D}}(x) = P(X > x | X \in \overline{D})$ since we don't have any information from the observed data about the proportion of observations that belong to the set $D$. Due to these identifiability problems it was assumed in Turnbull (1976) and Frydman (1994) that $P_S(D) = 0$. We need not assume in the sequel that $P_S(D)$ is equal to 0. The identifiability issues that arise from this assumption will be addressed later in Section 4. It is easy to see that $S_{\overline{D}}$ and $S$ give rise to the same likelihood. Therefore one should concentrate his efforts into finding a NPMLE of $S_{\overline{D}}$ when $P_S(D)$ is unknown. Let us denote the set $C$ as

$$C = \cup_{i=1}^{m}[q_i, p_i]$$

where $q_1 \leq p_1 < q_2 \leq p_2 < \ldots < q_m \leq p_m$. Let $s_j = S_{\overline{D}}(q_j{}^-) - S_{\overline{D}}(p_j{}^+)$. The likelihood given in (1) can be written as a function of $s_1, s_2, \ldots, s_m$ that is,

$$l(s_1, \ldots, s_m) = \prod_{i=1}^{n} \frac{\sum_{j=1}^{m} \mu_{ij} s_j}{\sum_{j=1}^{m} \nu_{ij} s_j} \tag{2}$$

where $\mu_{ij} = I_{[\ [q_j,p_j]\subset A_i]}$ and $\nu_{ij} = I_{[\ [q_j,p_j]\subset B_i]}$, $i = 1,\ldots,n$ and $j = 1,\ldots,m$. The NPMLE of $S_{\overline{D}}$ is actually not unique but there is a class of NPMLE's of $S_{\overline{D}}$ that share the same values $s_1, s_2, \ldots, s_m$ as it can be deduced by the following Lemma.

**Lemma 2** *For fixed values of $S_{\overline{D}}(q_j{}^-)$ and $S_{\overline{D}}(p_j{}^+)$, for $1 \le j \le m$, the likelihood is independent of how the decrease actually occurs in the interval $[q_j, p_j]$, so that $S_{\overline{D}}$ is undefined within each interval $[q_j, p_j]$.*

# 3 Nonproportional hazards models

The hazard rate of an individual with $p$-dimensional covariate vector $z$, for the proportional hazards model, is given as

$$h(t|z) = e^{\beta^T z} h_0(t)$$

where $\beta \in R^p$ is the parameter of interest and $h_0(t)$ is the baseline hazard rate. When a positive random variable $\eta$, called frailty, is introduced to act multiplicatively on the above hazard intensity function we obtain

$$h(t|z,\eta) = \eta e^{\beta^T z} h_0(t)$$

and equivalently,

$$S(t|z,\eta) = e^{-\eta e^{\beta^T z}\Lambda(t)}$$

where $\Lambda(t)$ is the baseline cumulative hazard function. Thus,

$$S(t|z) = \int_0^\infty e^{-x e^{\beta^T z}\Lambda(t)} dF_\eta(x) \equiv e^{-G(e^{\beta^T z}\Lambda(t))} \tag{3}$$

where

$$G(y) = -\ln(\int_0^\infty e^{-xy} dF_\eta(x))$$

and $F_\eta$ is the distribution function of the frailty parameter assumed in what follows to be completely known. When $G(x) = x$, the above model reduces to the Cox model. A well known frailty model is the Clayton-Cuzick model (Clayton and Cuzick 1985 and 1986) which corresponds to a Gamma distributed frailty.

The class of semiparametric transformation models as was defined in Cheng *et al.* (1995) for right-censored data, namely,

$$g(S(t|z)) = h(t) + \beta^T z$$

is equivalent to our class of models (3) through the relations

$$g(x) \equiv log(G^{-1}(-\log(x)), \quad h(t) \equiv \log(\Lambda(t))$$

where $g$ is known and $h$ unknown.

Let $(X_1, Z_1), ..., (X_n, Z_n)$ be i.i.d. random pairs of variables with marginal survival function defined in (3) as in Vonta (1996) and Slud and Vonta (2003). The function $G \in \mathcal{C}^3$ is assumed to be a known strictly increasing concave function with $G(0) = 0$ and $G(\infty) = \infty$. As in the previous section we assume that the random variables $X_i$ are incomplete due to arbitrary censoring and truncation. The likelihood (1) written for the frailty models defined in (3) takes the form

$$l(\Lambda, \beta|z) = \prod_{i=1}^{n} l_i(\Lambda, \beta|z) = \prod_{i=1}^{n} \frac{\sum_{j=1}^{k_i} \left\{ e^{-G(e^{\beta^T z} \Lambda(L_{ij}{}^-))} - e^{-G(e^{\beta^T z} \Lambda(R_{ij}{}^+))} \right\}}{\sum_{j=1}^{n_i} \left\{ e^{-G(e^{\beta^T z} \Lambda(\mathcal{L}_{ij}{}^+))} - e^{-G(e^{\beta^T z} \Lambda(\mathcal{R}_{ij}{}^-))} \right\}}. \tag{4}$$

Our goal is to obtain the joint NPMLE's of $\beta$, the parameter of interest and $\Lambda$, the nuisance parameter. In the maximization of (4) with respect to $\Lambda$ we employ Lemmas 1 and 2 that continue to hold under the present generalization with some adjustments. In particular, we give here Lemma 3, the proof of which retraces the steps of the corresponding Lemma 1 given in Turnbull (1976) and Alioum and Commenges (1996), as well as Lemma 4.

**Lemma 3** *Any cumulative hazard-type function $\Lambda$ within model (3) which increases outside the set $C \cup D$ cannot be the NPMLE of $\Lambda$.*

**Proof.**We will show first that any function $\tilde{\Lambda}$ which is not constant outside the set $Q$ cannot be the NPMLE of $\Lambda$. Define the points $r_j$ that belong to the interval $(p'_j, q'_{j+1})$, $1 \leq j \leq v-1$, where $r_j$ is some value greater than all the right and less than all the left endpoints in $[p'_j, q'_{j+1}]$. Let the function $\tilde{\Lambda}$ have jumps outside the set $Q$. There is at least one $r_k$, $1 \leq k \leq v-1$ for which either (i) $\tilde{\Lambda}(p'^+_k) < \tilde{\Lambda}(r_k) \leq \tilde{\Lambda}(q'^-_{k+1})$ or (ii) $\tilde{\Lambda}(p'^+_k) \leq \tilde{\Lambda}(r_k) < \tilde{\Lambda}(q'^-_{k+1})$ . Let $\Lambda^*$ be constant outside the set $Q$ and particularly $\Lambda^*(p'^+_k) = \Lambda^*(q'^-_{k+1}) = \tilde{\Lambda}(r_k)$ and $\Lambda^*(x) = \tilde{\Lambda}(x)$ for all $x \notin [q'_k, p'_{k+1}]$. Suppose that case (i) occurs. Then $\tilde{\Lambda}(p'^+_k) < \Lambda^*(p'^+_k)$ and consequently, since $G$ is an increasing function $e^{-G(e^{\beta^T z} \Lambda^*(p'^+_k))} < e^{-G(e^{\beta^T z} \tilde{\Lambda}(p'^+_k))}$. Because of the way the set $Q$ was constructed there is at least one observation $i$ such that $p'_k = R_{il}$ for $1 \leq l \leq k_i$ or $p'_k = \mathcal{L}_{il}$ for $1 \leq l \leq n_i$. Let $\mathcal{K}$ be the set of these observations. Then we have either

$$e^{-G(e^{\beta^T z \Lambda^*(R^+_{il})})} < e^{-G(e^{\beta^T z \tilde{\Lambda}(R^+_{il})})}$$

or

$$e^{-G(e^{\beta^T z \Lambda^*(\mathcal{L}^+_{il})})} < e^{-G(e^{\beta^T z \tilde{\Lambda}(\mathcal{L}^+_{il})})}.$$

It follows that $l_i(\Lambda^*, \beta|z) > l_i(\tilde{\Lambda}, \beta|z)$ for all $i \in \mathcal{K}$. For $i \notin \mathcal{K}$ we have that $l_i(\Lambda^*, \beta|z) \geq l_i(\tilde{\Lambda}, \beta|z)$. It is easy to see now that $l(\Lambda^*, \beta|z) > l(\tilde{\Lambda}, \beta|z)$, that is, the function $\tilde{\Lambda}$ cannot be the NPMLE of $\Lambda$ in likelihood (4). We obtain the same result in case (ii). This comment implies that for a $\hat{\Lambda}$ to be a suitable candidate for a NPMLE it has to be flat outside the set $Q$. Such a $\hat{\Lambda}$ is also flat in $W$. Therefore, the function $\hat{\Lambda}$ that maximizes likelihood (4) puts mass only in the set $C \cup D$ and remains flat outside this set. □.

**Lemma 4** *For fixed values of* $\Lambda(q_j{}^-)$ *and* $\Lambda(p_j{}^+)$*, for* $1 \le j \le m$*, the likelihood is independent of how the increase actually occurs in the interval* $[q_j, p_j]$*, so that* $\Lambda$ *is undefined within each interval* $[q_j, p_j]$*.*

We continue now to write the log-likelihood in the nonproportional hazards case in a more convenient form so that the maximization with respect to $\Lambda$ and $\beta$ will be possible. Since the set $C = \cup_{j=1}^{m}[q_j, p_j]$, the set $D$ can be written as $D = \cup_{j=0}^{m} D_j$, where $D_j = D \cap (p_j, q_{j+1})$, $p_0 = 0$ and $q_{m+1} = \infty$. Notice that $D_j$ is either a closed interval or a union of disjoint closed intervals. Let $\delta_j = P_\Lambda(D_j)$ denote the mass of the cumulative hazard function $\Lambda$ on the set $D_j$. From Lemma 3 we have that $\Lambda(q_j^-) = \Lambda(p_{j-1}^+) + \delta_{j-1}$ for $1 \le j \le m+1$. The log-likelihood can then be expressed as

$$\begin{aligned}\log l(\Lambda, \beta|z) &= \sum_{i=1}^{n} \Big\{ \log\Big( \sum_{j=1}^{m} \mu_{ij} \big( e^{-G(e^{\beta^T z}(\Lambda(p_{j-1}^+)+\delta_{j-1}))} - e^{-G(e^{\beta^T z}\Lambda(p_j{}^+))} \big) \Big) - \\ &\quad \log\Big( \sum_{j=1}^{m} \nu_{ij} \big( e^{-G(e^{\beta^T z}(\Lambda(p_{j-1}^+)+\delta_{j-1}))} - e^{-G(e^{\beta^T z}\Lambda(p_j{}^+))} \big) \Big) \Big\}. \qquad (5)\end{aligned}$$

In most real data problems, the set $D$ consists of the union of two intervals, namely, $D_0$ and $D_m$. If there are only right-truncated data involved then the set $D = D_m$. If there are only left-truncated data involved then the set $D = D_0$. Therefore the case $D = D_0 \cup D_m$ covers most of the problems one would encounter in practice and therefore we will deal with this case from now on as far as the examples are concerned. We will address the more general problem though from the point of view of the identifiability in Section 4. In the above special case we have $\delta_1 = \delta_2 = \ldots = \delta_{m-1} = 0$ and therefore likelihood (5) involves the parameters $\beta, \delta_0, \Lambda(p_0), \ldots, \Lambda(p_m)$. Since $\Lambda(p_0) = 0$ we have to maximize likelihood (5) with respect to the $p+m+1-$dimensional parameter $(\beta, \delta_0, \Lambda(p_1), \ldots, \Lambda(p_m))$. Notice that $\delta_m$ could be obtained directly from $\Lambda(p_m)$. Similarly to Finkelstein *et al.* (1993) and Alioum and Commenges (1996) we will make the reparametrization $\gamma_0 = \log(\delta_0)$ and $\gamma_j = \log(\Lambda(p_j))$ for $j = 1, \ldots, m$ for computational convenience. Therefore the log-likelihood becomes

$$\begin{aligned}\log l(\Lambda, \beta|z) &= \sum_{i=1}^{n} \Big\{ \log\Big( \sum_{j=1}^{m} \mu_{ij} \big( e^{-G(e^{\beta^T z + \gamma_{j-1}})} - e^{-G(e^{\beta^T z + \gamma_j})} \big) \Big) - \\ &\quad \log\Big( \sum_{j=1}^{m} \nu_{ij} \big( e^{-G(e^{\beta^T z + \gamma_{j-1}})} - e^{-G(e^{\beta^T z + \gamma_j})} \big) \Big) \Big\}. \qquad (6)\end{aligned}$$

A second reparametrization which ensures monotonicity of the sequence $\gamma_j$ was subsequently employed, that is, $\tau_1 = \gamma_1$ and $\tau_j = \log(\gamma_j - \gamma_{j-1})$ for $j = 2, \ldots, m$. This parametrization improved also the speed of the convergence. The maximization in section 5 was actually done with respect to the parameters $\beta$ and $\gamma_0, \tau_1 \ldots, \tau_m$ with the use of software such as Splus and Fortran 77.

# 4 Identifiability

In our discussion of the identifiability of $\Lambda$ and $\beta$ we have to examine two cases, namely, the case $\beta = 0$ and the case $\beta \neq 0$ and comment on each of them separately. For the case where there are no covariates,

that is, when $\beta = 0$, the cumulative hazard function $\Lambda$ is not identifiable. In order to show this we will concentrate on the case where $D = D_0 \cup D_m$ which is general enough as we argued in the previous section. Let us define the family of cumulative hazard functions indexed by two positive constants $c_1$ and $c_2$ as follows

$$\ell(t, c_1, c_2) = G^{-1}(c_1 + \min(\Lambda(t), c_2))$$

for $t \in \overline{D}$. This class of cumulative hazard functions gives rise to the same likelihood as $\Lambda$ for any value of the constant $c_1$ and for the constant $c_2$ taken large enough. For an individual $i$, in the simple case where $k_i = n_i = 1$, the $i^{\text{th}}$ term of the likelihood for the family $\ell$ of cumulative hazard functions and for $\beta = 0$ is given as

$$l_i(\ell, 0|z) = \frac{e^{-G(G^{-1}(c_1+\min(\Lambda(L_i{}^-),c_2))} - e^{-G(G^{-1}(c_1+\min(\Lambda(R_i{}^+),c_2))}}{e^{-G(G^{-1}(c_1+\min(\Lambda(\mathcal{L}_i{}^+),c_2))} - e^{-G(G^{-1}(c_1+\min(\Lambda(\mathcal{R}_i{}^-),c_2))}}$$

$$= \frac{e^{-\min(\Lambda(L_i{}^-),c_2)} - e^{-\min(\Lambda(R_i{}^+),c_2)}}{e^{-\min(\Lambda(\mathcal{L}_i{}^+),c_2)} - e^{-\min(\Lambda(\mathcal{R}_i{}^-),c_2)}}$$

which is equal to $l_i(\Lambda, 0|z)$ for any $c_1$ and $c_2$ chosen larger than the largest right endpoint $p_m$ in $C$.

For the case $\beta \neq 0$ the identifiability argument depends heavily on our assumption of a frailty model. Let us concentrate first in the case where the set $D = D_0 \cup D_m$ where $D_0 = [0, d_0]$ and $D_m = [d_m, \infty)$. We will prove that when we have at least two covariates then we can identify the parameter $\beta$ along with the parameters $(\delta_0, \Lambda(p_1), \ldots, \Lambda(p_m), \delta_m)$. In particular, in order to show the identifiability of $\beta$ and $\Lambda$ we show that they are both functions of quantities that are known to be identifiable. For convenience let us assume that the two covariates $z_1$ and $z_2$ are binary, giving rise to four combinations of observations. Following the construction of the set $Q$ presented in Section 2 for each of the four combinations separately we produce four sets of the type $C \cup D$. We denote by $C^{00} \cup D^{00}$ the set that corresponds to the observations with $z_1 = z_2 = 0$, by $C^{10} \cup D^{10}$ the observations with $z_1 = 1$ and $z_2 = 0$ and similarly for the other two groups. Then $D^{00} = D_0^{00} \cup D_m^{00}$ and moreover, $D_0^{00} = [0, d_0^{00}]$ and $D_m^{00} = [d_m^{00}, \infty)$ while similar notations hold for the other three groups. Let $u_0^\star = \max\{d_0^{00}, d_0^{10}, d_0^{01}, d_0^{11}\}$, $u_m^\star = \min\{d_m^{00}, d_m^{10}, d_m^{01}, d_m^{11}\}$ and $U = [u_0^\star, u_m^\star]$. Let also $C^\star = C^{00} \cap C^{10} \cap C^{01} \cap C^{11} = \cup_{i=1}^{m'}[q_i^\star, p_i^\star]$.

The quantities

$$S_U(p_j^\star|z) = \frac{S(p_j^\star|z) - S(u_m^\star|z)}{S(u_0^\star|z) - S(u_m^\star|z)} \tag{7}$$

for $(z_1, z_2)$ equal to (0,0) or (0,1) or (1,0) or (1,1) and for $j = 1, \ldots, m'$ are identifiable (Lagakos *et al.* (1988), Finkelstein *et al.* (1993)). We assume here that the mass in the interval $[q_j^\star, p_j^\star]$ is concentrated at the point $p_j^\star$ since we have no way of knowing how exactly is that mass distributed in that interval. Another identifiable quantity that is available is the ratio of the hazard functions $h_U(x|z) = (-\log S_U(x|z))'$ for two different values of $z$, taken at $x = p_j^\star$. This quantity is equal to

$$H_U(x|z_l, z_k) = \frac{h_U(x|z = z_l)}{h_U(x|z = z_k)}$$

$$= \frac{e^{-G(e^{\beta^T z_k}\Lambda(x))} - e^{-G(e^{\beta^T z_k}\Lambda(u_m^\star))}}{e^{-G(e^{\beta^T z_l}\Lambda(x))} - e^{-G(e^{\beta^T z_l}\Lambda(u_m^\star))}} e^{-G(e^{\beta^T z_l}\Lambda(x)) + G(e^{\beta^T z_k}\Lambda(x))} e^{\beta} \frac{G'(e^{\beta^T z_l}\Lambda(x))}{G'(e^{\beta^T z_k}\Lambda(x))}. \tag{8}$$

So, from (7), $\Lambda(p_j^\star)$ can be obtained as a function $f$ of $\Lambda(u_0^\star)$, $\Lambda(u_m^\star)$ and the identifiable quantity $S_U(p_j^\star|0,0)$. Then, $\Lambda(u_m^\star)$ can be obtained as function $f_1'$ of $\beta$, $\Lambda(u_0^\star)$, the function $f$ and the identifiable quantity $S_U(p_j^\star|0,1)$. In other words, $\Lambda(u_m^\star)$ can be obtained as a function $f_1$ of $\beta$, $\Lambda(u_0^\star)$, $S_U(p_j^\star|0,0)$ and $S_U(p_j^\star|0,1)$. Then similarly, $\Lambda(u_0^\star)$ can be obtained as a function $f_2$ of $\beta$ and the quantities $S_U(p_j^\star|0,0)$, $S_U(p_j^\star|0,1)$ and $S_U(p_j^\star|1,0)$. Consequently, $\beta_1$ the first component of the vector $\beta$, can be obtained as a function $f_3$ of the quantities $S_U(p_j^\star|0,0)$, $S_U(p_j^\star|0,1)$, $S_U(p_j^\star|1,0)$ and $S_U(p_j^\star|1,1)$ and $\beta_2$, the second component of the vector $\beta$. Finally, $\beta_2$ is identifiable since it can be obtained as a function $f_4$ of the identifiable quantities $S_U(p_j^\star|0,0)$, $S_U(p_j^\star|0,1)$, $S_U(p_j^\star|1,0)$, $S_U(p_j^\star|1,1)$ and $H_U(p_j^\star|z_l,z_k)$ for some choice of values $z_l$ and $z_k$. Having identified $\beta_2$ we follow this argument backwards to obtain identifiability of $\beta_1$ and $\Lambda(p_j^\star)$ for $j=1,\ldots,m'$.

We need to show now identifiability at all the points $p_j$, $j=1,\ldots,m$, $d_0$, and $d_m$. We show identifiability at $d_m$ and similar discussions for the other points will complete the argument. Note that there exists a value $z_l$ of $z$ for which $d_m^{z_l} = d_m$. Then the quantity

$$\frac{S(u_m^\star|z=z_l) - S(d_m|z=z_l)}{S(u_0^\star|z=z_l) - S(d_m|z=z_l)}$$

is identifiable. Therefore, $\Lambda(d_m)$ is identifiable as a function of the preceding identifiable quantity and the identifiable quantities $\beta$, $\Lambda(u_0^\star)$ and $\Lambda(u_m^\star)$.

It is easy to see that in the above situation, we can have identifiability even when only one covariate, with takes at least three different values, is present. On the other hand, if we have only one covariate which takes two values then we could not identify the mass on $D_0$ and $D_m$ separately. We could however identify the global mass on $D$ acting as though all the mass is concentrated on $D_0$, which is the case of left-truncated data, or on $D_m$ which is the case of right-truncated data.

In more general truncating schemes where $D = \cup_{j=0}^m D_j$ we can obtain identifiability of all the quantities involved, by looking at windows in time that follow the pattern that we have already examined, namely, a set $D$ followed by a set $C$ and then followed by a set $D$. We could for example start at the last window in time that follows this pattern and the window next to the last. Let us denote the time interval corresponding to the last window, $U_m$ and to the window next to the last, $U_{m-1}$. Conditionally on the fact that we are in $U_m$, we follow the argument described above to identify $\beta$, $\delta_{m-1}$, $\delta_m$ and $\Lambda(p_j)$ for those $p_j$ that fall between the sets $D_{m-1}$ and $D_m$. Similarly, conditionally on the fact that we are in $U_{m-1}$, we can identify the quantities $\beta$, $\delta_{m-2}$, $\delta_{m-1}$ and $\Lambda(p_j)$ for those $p_j$ that fall between the sets $D_{m-2}$ and $D_{m-1}$. Since we have identified the mass at $D_{m-1}$ conditionally belonging to $U_m$ and conditionally belonging to $U_{m-1}$, we can easily find its 'real' mass when it belongs to $U_{m-1} \cup U_m$. Using this value we can adjust for all the quantities involved in the new time interval $U_{m-1} \cup U_m$. We continue in the same way, namely, looking at the window $U_{m-2}$ and the time interval $U_{m-1} \cup U_m$, which share mass at $D_{m-2}$, retracing the same steps until we identify all the quantities involved in the time interval

$U_{m-2} \cup U_{m-1} \cup U_m$. The procedure continues until all the windows down to the last one denoted by $U_1$ have been covered.

## 4.1 Real data

In this section our model is illustrated with a previously analyzed data set (Kalbfleisch and Lawless (1989), Lagakos *et al.* (1988) and Alioum and Commenges (1996)) on tranfusion related AIDS. The data set consists of 494 cases, of which only 295 are consistent in the sense that their infection could be attributed to a single transfusion or to a short series of transfusions. Our analysis is based on those 295 cases diagnosed by June 30, 1986 and reported to the Centers for Disease Control in Atlanta, Georgia prior to January 1, 1987. For each individual the time of infection $x_i$ (in months), the induction period $t_i$ (also in months), and the age+1 years at the time of transfusion are reported. The earliest infection was reported in January 1978 and labeled as month 1 so that the maximum observable induction time is $x_*$ equal to 102 months. The data are right-truncated because an individual $i$ is only included in the sample if $x_i + t_i \leq x_*$. We consider three groups of individuals according to their age, namely children, adults and elderly (with corresponding age intervals [0,12], (12,60) and [60,80]) and create a covariate with levels 0, 1 and 2. This partition appears in Kalbfleisch and Lawless (1989). As was shown in Section 4, the presence of this covariate allows our model to be identifiable. In our analysis we keep the month as the unit of time (as in Alioum and Commenges (1996)). Specifically, $x_i$ is shifted to $x_i - 0.5$ (as in Kalbfleisch and Lawless (1989) for their continuous analyses of the data) and $t_i$ is shifted to $t_i + 0.5$ and assumed to lie in a censoring interval equal to $[t_i, t_i + 1)$.

Two frailty models of the class defined in (3) are considered, namely, the Inverse Gaussian and the Clayton-Cuzick. For the first model, $\eta$ is taken to be distributed as Inverse Gaussian with mean 1 and variance $1/2b$. For the Clayton-Cuzick model, $\eta$ is taken to be distributed as Gamma with mean 1 and variance $c$. When one chooses a value for the parameter $b$ or $c$, one actually specifies the variance of the frailty. The function $G$ takes respectively the form

$$G(x, b) = \sqrt{4b(b + x)} - 2b, \ b > 0$$

and

$$G(x, c) = \frac{1}{c} \ln(1 + cx), \ c > 0.$$

In Table 1 we present the estimator of $\beta$ and $\widehat{P_\Lambda(D)}$, that is, the estimated probability of being truncated, as obtained by the maximization of our likelihood with respect to $\beta$ and $\Lambda$, along with the log-likelihood value at $(\hat{\beta}, \hat{\Lambda})$ for different values of the parameter $b$ of the Inverse Gaussian model. In Table 2 the corresponding quantities for the Clayton-Cuzick model and for different values of the parameter $c$ are presented. In both cases we maximized over a total of 72 parameters, as the value of $m$, defined in Section 2, was found to be equal to 71.

HERE : Table 1. Inverse Gaussian Model.

HERE : Table 2. Clayton-Cuzick Model.

Notice that for the Inverse Gaussian model the maximum of the log-likelihood with respect to the parameter $b$ occurs at $b = 0.039$ providing a $\hat{\beta} = -4.69$ and $\widehat{P_\Lambda(D)} = 0.82$. On the other hand, for the Clayton-Cuzick model the log-likelihood values continue to increase as $c$ tends to $\infty$. However they remain quite stable after $c = 10$, for which $\hat{\beta} = -3.62$ and $\widehat{P_\Lambda(D)} = 0.88$.

Figures 1, 2 and 3 present the fit of the Inverse Gaussian model at $\hat{b} = 0.039$ to the Aids data for the three age groups separately. Each figure provides the estimated survival function through the Inverse Gaussian frailty model as well as the corresponding nonparametric maximum likelihood estimator of the survival function (as in Turnbull (1976)). Figures 4, 5 and 6 present similarly the fit of the Clayton-Cuzick model with $c = 10$ to the Aids data for the same three age groups. The nonparametric maximum likelihood estimator of the survival function for each group, presented in the figures, was found as the maximizer of the likelihood defined in (2) with respect to the vector $(s_1, \ldots, s_m)$, where the value of $m$ is of course different for each group. The maximization was done subject to the constraints $\sum s_j = 1$ and $s_j \geq 0$ for $1 \leq j \leq m$. Note that the self-consistency algorithm proposed by Turnbull (1976) was used in the process. Recall that the survival function $S$ is not identifiable unless $P_S(D)$ is known and this is due to the fact that the likelihood (2) does not depend on $P_S(D)$, the probability of belonging to the set $D$. Therefore in order to be able to compare our results with the NPMLE's we have matched $P_S(D)$ with the probability $\widehat{P_\Lambda(D)}$ obtained through the maximization of our log-likelihood (3).

Observe that the best fit to the Aids data is obtained through the Inverse Gaussian model, which becomes evident (Figure 7) from the first age group, namely, the children less than 12 years of age. Since the fit of the Inverse Gaussian frailty model is substantially better as opposed to the fit due to the proportional hazards model we can safely deduce that a random effect is essential to be included in the hazard rate in order to describe the heterogeneity present in the transfusion related Aids data. This conclusion is greatly supported also by the fact that the log-likelihood in the case of the Inverse Gaussian model is maximized at $b = 0.039$ implying a big variance for the frailty parameter equal to 12.82. At the same time, in the Clayton-Cuzick model, the log-likelihood values increase as the parameter $c$, which is actually the variance of the frailty parameter, tends to $\infty$, although not by much after $c$ equal to 10. Observe also that the fit of the Inverse Gaussian model with frailty variance equal to 12.82 is better than the fit of the Clayton-Cuzick model with about the same variance as it can be seen mainly when comparing Figures 1 and 4 that correspond to the children's group.

HERE: FIGURES 1 to 7

## 4.2 Simulated data

We performed four simulations with 100 replications each. For each sample 400 times $x_i$ and 400 survival times $t_i$ were generated. Given a time $x_*$ the times $x_i$ were generated from a $U(0, x_*)$ distribution while the survival times $t_i$ were generated from a frailty model of the class defined in (3) with Weibull baseline hazard function. More specifically, we considered a Weibull distribution with scale parameter $\rho_0$ equal to 2 and shape parameter $\kappa$ equal to 0.7, where the baseline cumulative hazard function is of the form $\Lambda(t) = (t/\rho_0)^\kappa$. Two binary covariates $Z_1$ and $Z_2$ for which $P(Z_i = 0) = P(Z_i = 1) = 1/2$ for $i = 1, 2$ were considered. From the data that were generated only the data that satisfied the condition $x_i + t_i \leq x_*$ were kept in the sample giving rise to right-truncated data. The interval $[0, x_*]$ was divided into $n$ =15 equal intervals that constitute a partition $[a_{k-1}, a_k)$ for $k = 1, \ldots, n$ where $a_0 = 0$ and $a_n = x_*^+$. The survival times $t_i$ are not only truncated but also interval-censored as they are reported only to belong in one of those intervals of the partition. In fact, they are reported to belong to intervals of the form $[a_{k-1}, x_* - x_i)$ whenever happens that $x_* - x_i < a_k$. Two of the simulations were generated from the Clayton-Cuzick model with parameter $c$ equal to 2 and 0.5 respectively and two from the Inverse Gaussian model with parameter $b$ equal to 0.5 and 1 respectively. The true probability of belonging to the set $D$, $P_\Lambda(D)$ was taken to be 0.19 (as was used for example in Finkelstein *et al.* (1993)) for all simulations. Because of truncation, the sample size of the generated samples is random and so is the number of parameters to be estimated. The sample size was about 300 and the number of parameters to be estimated for each sample about 30. Also, the point $x_*$ varied according to the true values of $\beta_1$, $\beta_2$ and $P_\Lambda(D)$.

In the next tables we provide the mean of the estimators $\hat{\beta}_1$, $\hat{\beta}_2$ and $\widehat{P_\Lambda(D)}$ as well as their sample variances for different true values of $\beta_1$, $\beta_2$ and $b$ or $c$ involved in the frailty distribution. $\widehat{P_\Lambda(D)}$ was obtained from the overall population in each sample, as a weighted average of the survival curves of the four groups at the point $p_m$.

HERE : Tables 1 to 4

In our simulations we were faced with the situation of the likelihood having a saddle point and therefore our procedure of maximization diverged. This small proportion of samples was left out as we did not use a sophisticated enough method of maximization as other authors (Alioum and Commenges (1996), Pan and Chappell (2002)). We found that the estimator of the regression coefficient which is bigger in absolute value, tends to slightly overestimate its true absolute value resulting also in higher variance than the estimator of the other coefficient which in general behaves well. Therefore the median will be a more robust estimator in this case. More specifically, the median of $\beta_1$ is 2.109 for Table 1, 2.193 for Table 2, 2.087 for Table 3 and -2.127 for Table 4. The median of $\beta_2$ is -1.235 for Table 1, -1.289 for Table 2, -1.244 for Table 3 and -1.069 for Table 4. The mean of $\widehat{P_\Lambda(D)}$ seems to be a very good estimator of $P_\Lambda(D)$ in all situations. There is a tendency however for both frailty models to underestimate $P_\Lambda(D)$ as the

baseline survival function appears to be underestimated as well, although more work is needed to draw safely such a conclusion.

# 5 Discussion

In this paper we are dealing with the most general scheme of truncated and censored data in survival analysis, as in Alioum and Commenges (1996). In their paper, Alioum and Commenges are considering a Cox model, while, under the same pattern of data, we are dealing with a generalization of this model in order to take into account a possible heterogeneity among the population. Using a likelihood, along the lines of Turnbull (1976), Finkelstein *et al.* (1993), Alioum and Commenges (1996), Pan and Chappell (2002), we obtain nonparametric estimates for the underlying baseline cumulative hazard, the coefficients of the underlying Cox regression, and the probability of truncation.

The set of real data on AIDS acquired by transfusion, previously analyzed by many authors, was for the first time analyzed under the framework of the frailty models. It was a very interesting finding that heterogeneity is in fact present in the AIDS data and as we argued in the previous section, it was very well described by the Inverse Gaussian distribution. On the other hand, we feel that the Gamma frailty model is not appropriate for this set of data for many reasons. One reason is that the relative heterogeneity among the patients seems to reduce as time passes by, instead of remaining constant as is the case for the Gamma frailty (Hougaard (1984)). It is generally known that the Gamma model, although popular, has many drawbacks one of which as Sahu and Dey (2003) point out, is that it weakens the effect of the covariates. This is exactly what we came to realize through our work with the Gamma model. This feature becomes apparent when one takes a closer look not only at the truncation set overall estimated probability but also at the same estimated probability for each group (see Figures 1 through 6). Notice also that the maximization of the likelihood with respect to the parameter $c$ tending to $\infty$, implies an estimated probability of truncation tending to 1 which is unrealistic.

Our estimated right-truncation probability under the Inverse Gaussian model, that is, $\widehat{P_\Lambda(D)} = 0.82$ seems also to be rather high. But in view of the values for $P_\Lambda(D)$ that were assumed "a priori" by former authors, (see for instance Alioum and Commenges who thought it very plausible that it would be equal to 0.60), and also taking into account the relatively good estimation of $P_\Lambda(D)$ we obtain in the simulations, we think that this is a reliable value for $P_\Lambda(D)$. Another reason for comforting us in our opinion that the Inverse Gaussian model provides a good fit to the AIDS data, is the good fit we obtain in each group for the survival obtained by our model as compared to the respective Turnbull conditional NPMLE of the survival. Finally, the frailty models considered in this paper, are part of a more general class of transformation models (see Bagdonavicius and Nikulin (2002)), so that our method could be generalized to this class of models under regularity conditions on the transformation function $G$.

# REFERENCES


Alioum, A. and Commenges, D. (1996). A proportional hazards model for arbitrarily censored and truncated data. *Biometrics* **52**, 512-524.

Bagdonavicius, V. and Nikulin, M. (2002). *Accelerated Life Models*, Chapman and Hall/CRC, Boca Raton.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835-845.

Clayton, D. and Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *J. Roy. Statist. Soc.* **A 148**, 82-117.

Clayton, D. and Cuzick, J. (1986). The semiparametric Pareto model for regression analysis of survival times. *Papers on Semiparametric Models* MS-R8614, 19-31. Centrum voor Wiskunde en Informatica, Amsterdam.

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

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics***42**, 845-854.

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

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.

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

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

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387-396.

Huang, J. (1994). Efficient estimation for the Cox model with interval censoring. *Technical Report,* Department of Statistics, University of Washington.

Huang, J. and Wellner, J. A. (1995). Efficient estimation for the proportional model with "case 2" interval censoring. *Technical Report,* Department of Statistics, University of Washington.

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.

Lagakos, S. W., Barraj, L. M. and De Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika* **75**, 515-523.

Pan, W. and Chappell, R. (2002). Estimation in the Cox proportional hazards model with left-truncated and interval-censored data. *Biometrics* **58**, 64-70.

Sahu, S. K. and Dey, D. K. (2003). On multivariate survival models with a skewed frailty and a correlated baseline hazard process. *http://www.maths.soton.ac.uk/staff/Sahu/research/papers/logskew.html*.

Slud, E. V., and Vonta, F. (2003). Consistency of the NPML estimator in the right-censored transformation model. *Scandinavian Journal of Statistics* (to appear).

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

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. I. (1983). The deviant dynamics of death in heterogeneous populations. *RR-83-1* Laxenburg, Austria: International Institute for Applied Systems Analysis.

Vonta, F. (1996). Efficient estimation in a non-proportional hazards model in survival analysis. *Scandinavian Journal of Statistics* **23**, 49-61.

**Tables for section on Real data:**

**Table1 - Inverse Gaussian Model**

| $b$ | $\hat{\beta}$ | $\widehat{P_\Lambda(D)}$ | $LogLik$ |
|---|---|---|---|
| 0.005 | -5.961 | 0.939 | -992.8312 |
| 0.01 | -5.793 | 0.902 | -992.6381 |
| 0.02 | -5.302 | 0.864 | -992.4520 |
| 0.038 | -4.715 | 0.823 | -992.3735 |
| 0.039 | -4.690 | 0.821 | -992.3734 |
| 0.04 | -4.466 | 0.819 | -992.3736 |
| 0.05 | -4.454 | 0.802 | -992.3884 |
| 0.1 | -3.812 | 0.735 | -992.6017 |
| 0.5 | -2.723 | 0.553 | -994.0544 |
| 3.0 | -2.257 | 0.486 | -995.7565 |

| **Table 2 - Clayton-Cuzick Model** | | | |
|---|---|---|---|
| $c$ | $\hat{\beta}$ | $\widehat{P_\Lambda(D)}$ | $LogLik$ |
| 0.5 | -2.515 | 0.498 | -994.571 |
| 0.7 | -2.627 | 0.518 | -994.156 |
| 1.0 | -2.768 | 0.553 | -993.728 |
| 3.0 | -3.258 | 0.718 | -992.770 |
| 8.0 | -3.576 | 0.856 | -992.417 |
| 10.0 | -3.623 | 0.880 | -992.377 |
| 20.0 | -3.725 | 0.934 | -992.300 |
| 25.0 | -3.747 | 0.946 | -992.286 |
| 30.0 | -3.762 | 0.954 | -992.276 |
| 1000.0 | -3.836 | 0.998 | -992.232 |

**Tables for section on simulated data**

| **Table 1 - Clayton-Cuzick Model** | | |
|---|---|---|
| $\beta_1 = 2,\ \beta_2 = -1,\ c = 0.5$ | $Mean$ | $Sample\ Variance$ |
| $\hat{\beta}_1$ | 2.237 | 0.659 |
| $\hat{\beta}_2$ | -1.239 | 0.100 |
| $\widehat{P_\Lambda(D)}$ | 0.200 | 0.027 |

| **Table 2 - Clayton-Cuzick Model** | | |
|---|---|---|
| $\beta_1 = 2,\ \beta_2 = -1,\ c = 2.0$ | $Mean$ | $Sample\ Variance$ |
| $\hat{\beta}_1$ | 2.331 | 0.726 |
| $\hat{\beta}_2$ | -1.368 | 0.521 |
| $\widehat{P_\Lambda(D)}$ | 0.211 | 0.060 |

| **Table 3 - Inverse Gaussian Model** | | |
|---|---|---|
| $\beta_1 = 2,\ \beta_2 = -1,\ b = 1.0$ | $Mean$ | $Sample\ Variance$ |
| $\hat{\beta}_1$ | 2.269 | 0.617 |
| $\hat{\beta}_2$ | -1.270 | 0.101 |
| $\widehat{P_\Lambda(D)}$ | 0.204 | 0.029 |

| **Table 4 - Inverse Gaussian Model** | | |
|---|---|---|
| $\beta_1 = -2,\ \beta_2 = -1,\ b = 0.5$ | $Mean$ | $Sample\ Variance$ |
| $\hat{\beta}_1$ | -2.205 | 0.603 |
| $\hat{\beta}_2$ | -1.097 | 0.189 |
| $\widehat{P_\Lambda(D)}$ | 0.175 | 0.031 |

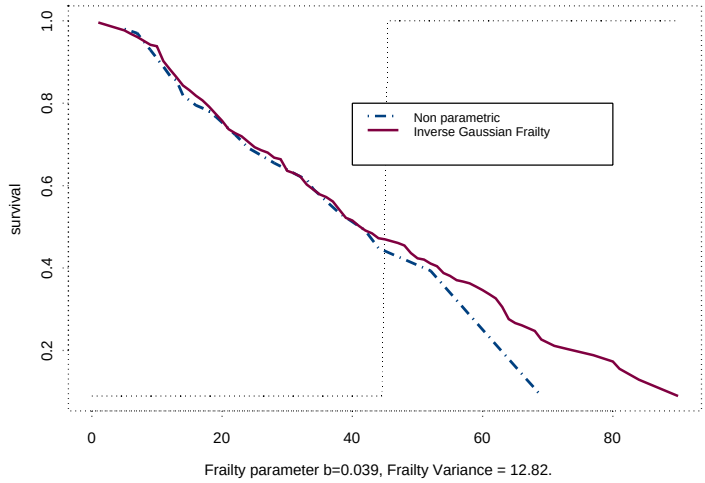


Figure 1: Comparison of survival curves estimates for children (group Z = 0).

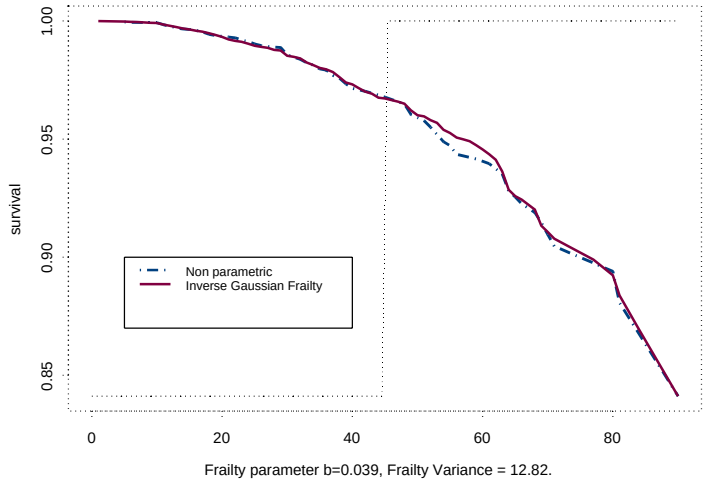


Figure 2: Comparison of survival curves estimates for adults (group Z = 1).

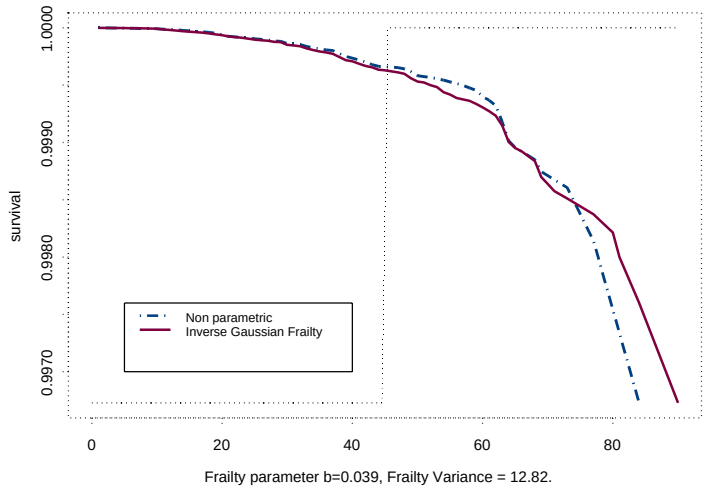


Figure 3: Comparison of survival curves estimates for elderly (group Z = 2).

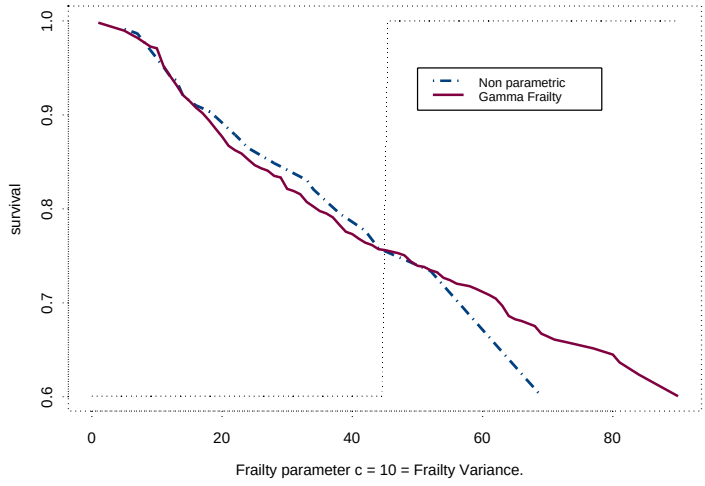


Figure 4: Comparison of survival curves estimates for children (group Z = 0).

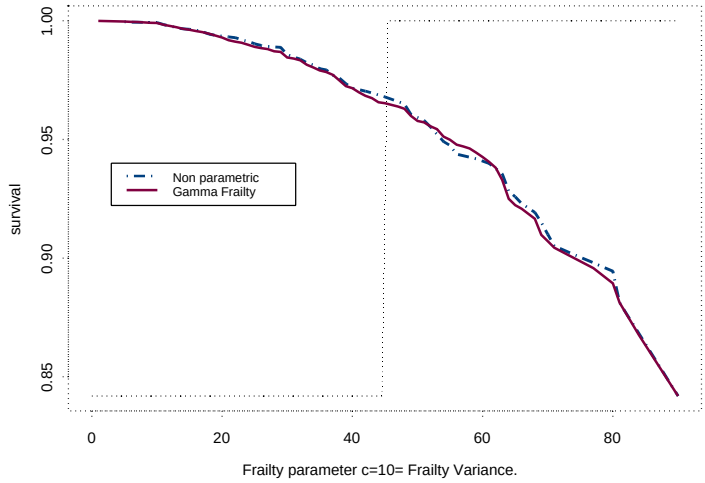


Figure 5: Comparison of survival curves estimates for adults (group Z = 1).

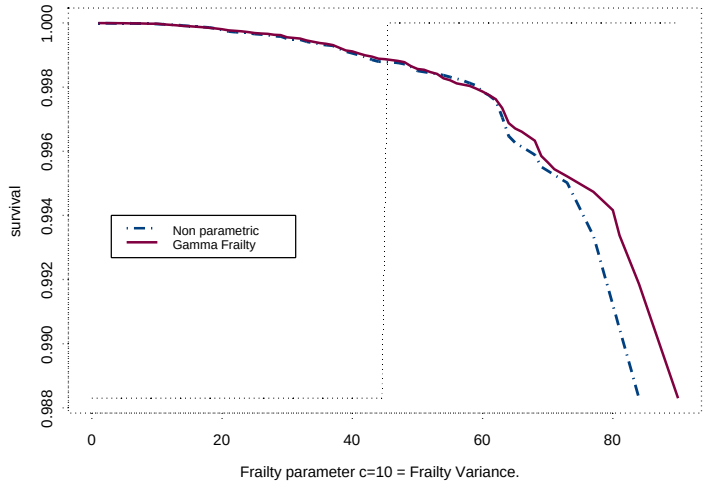


Figure 6: Comparison of survival curves estimates for elderly (group Z = 2).

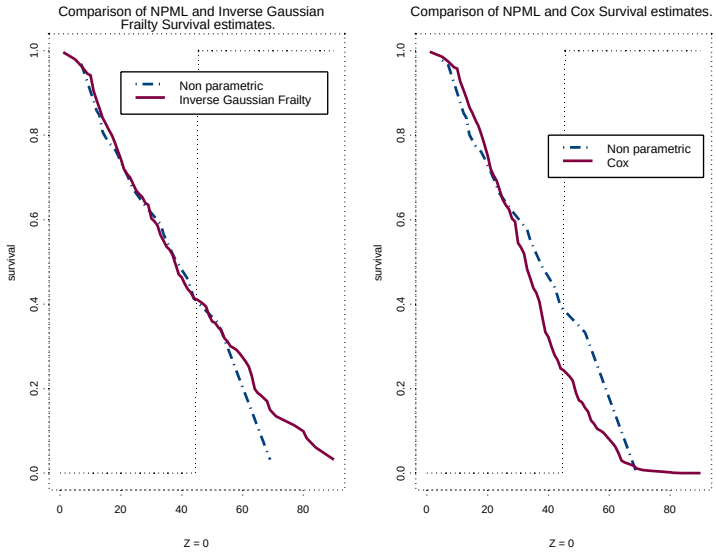


Figure 7: Comparison of survival curves estimates for children (group Z = 0, frailty b=0.039).

*Key-words*:
Frailty models; transformation models; censored data; truncated data; nonparametric maximum likelihood estimation; gamma frailty; inverse gaussian frailty.

*Corresponding author*:

Professeur C. Huber
Université René Descartes, UFR Biomédicale
45 rue des Saints-Pères
75270 Paris Cedex 06
tél. : 01-42-86-21-01
Fax : 01-42-86-04-02
E-mail: huber@univ-paris5.fr

*Affiliation of the authors*:

[a] MAP 5, FRE CNRS 2428, Université René Descartes, Paris 5, 45 rue des Saints-Pères, 75 006 Paris and U472 INSERM, 16 av. P-V Couturier, 94800 Villejuif, France.

[b] Department of Mathematics and Statistics, University of Cyprus P.O. Box 20537, CY-1678, Nicosia, Cyprus. E-mail: vonta@ucy.ac.cy.