Brazilian Journal of Probability and Statistics
2015, Vol. 29, No. 1, 132–144
DOI: 10.1214/13-BJPS232

The generalized time-dependent logistic frailty model: An application to a population-based prospective study of incident cases of lung cancer diagnosed in Northern Ireland

Eder A. Milani[a], Vera L. D. Tomazella[a], Teresa C. M. Dias[a] and Francisco Louzada[b]

[a]*Universidade Federal de São Carlos*
[b]*Universidade de São Paulo*

Abstract. Survival models with univariate frailty may be used when there is no information on covariates that are important to explain the failure time. The lack of information may be with respect to covariates that were not observed or even covariates which for some reason we can not measure, for instance, environmental or genetic factors. In this paper, we extend the generalized time-dependent logistic model proposed by Mackenzie (*The Statistician* **45** (1996) 21–34), by including a frailty term in the modeling. The proposed methodology uses the Laplace transform to find the survival function unconditional on the individual frailty. A simulation study examines the bias, the mean squared errors and the coverage probabilities. Estimation is based on maximum likelihood. A real example on lung cancer illustrates the applicability of the methodology, which is compared to the modeling without frailty via selection criteria to determine which model best fits the data.

1 Introduction

To express the distribution of an nonnegative random variable, T, which, in general, represents the random behavior of lifetime of individuals (or components) in some population, several mathematically equivalent functions that uniquely determine the distribution can be considered, namely the cumulative distribution, density, survival and hazard functions. The hazard function is particularly useful because of its interpretation as the way in which the instantaneous probability of failure of an individual changes with time. The seminal paper by Cox (1972) on hazard modeling opened a new era of statistical methodology for studying time-to-event data. His model advocates that the ratio of the failure rates of any two individuals are proportional. A strong assumption that may not be in accordance with several practical real situations. This fact has been determinant in the developing of several types of non-proportional hazard models. Among them we mention the accelerated failure model (Prentice, 1978), the hybrid hazard model (Etezadi-Amoli and Ciampi, 1987) and the extended hybrid hazard models (Louzada-Neto,

Key words and phrases. Generalized time-dependent logistic model, survival models with univariate frailty, non-proportional hazard model.
Received June 2011; accepted October 2013.

1997, 1999). In this paper, we focus on a parametric family of non-proportional hazard model, the so called generalized time-dependent logistic (GTDL) model (Mackenzie, 1996), which was proposed as a parametric competitor for the proportional hazard model.

Another possible peculiarity of the survival modeling mentioned above is related to the usual assumption made in analyzing the treatment effect, or hazard factors, in which individuals are conditionally independents given the observed covariates. When the data arise from multiple events or repeated events for the same individual, the assumption of independence can be questioned. Motivated by situations where there is possible dependence between individuals or groups of individuals Clayton (1978) and Oakes (1982) considered the first model for multivariate frailty. The frailty model is characterized by using a random effect, that is, a random variable which is not observed, or that represents information that is not or cannot be observed, but which affects the time of failure. When we only have one information for an individual, the random effect is introduced into the hazard function in order to control unobserved heterogeneity of units under study. For some studies with univariate frailty models interested readers can refer to Tomazella et al. (2008), which considered the Cox's model with frailty and under a Bayesian perspective considering reference prior distributions, Aalen and Tretli (1999), which applied the compound-Poisson distribution to data from testicular cancer and Henderson and Oman (1999), which studied the consequence of ignoring the frailty in the fitting.

In this paper, we envisage a scenario where we have only one observation per individual. Our main purpose is to extend the GTDL model introducing a frailty term. Our methodology uses the population (or unconditional) survival function. The Laplace transform of the frailty density is instrumental to obtain such a function.

The paper is organized as follows. Section 2 presents the GTDL and GTDL frailty models, as well as the construction of the likelihood for the model and the respective estimation procedure. A simulation study which examines the bias, the mean squared errors and the coverage probabilities is presented in Section 3. In Section 4, a real example on lung cancer illustrates the applicability of the methodology, which is compared to the modeling without frailty via selection criteria to determine which model best fits the data. Some final comments are in the Section 5.

2 Model formulation

In this section, we introduce the GTDL regression model and its frailty version, presenting the hazard, the survival, the probability density functions and the particular case when the GTDL model and GTDL with fralty model become a long

duration model (Ibrahim et al., 2001), that is, a model which considers that a proportion of the population is cured or not susceptible to the event of interest. Also, we present the inferential approach.

2.1 GTDL regression model

The GTDL regression model, which is a non-proportional hazard model, is defined by the hazard function, given by

$$h(t|\lambda,\alpha,\boldsymbol{\beta}) = \lambda \frac{\exp(\alpha t + \mathbf{x}'\boldsymbol{\beta})}{1+\exp(\alpha t + \mathbf{x}'\boldsymbol{\beta})}, \tag{2.1}$$

where $\lambda > 0$ is a scalar, α is a measure of the time effect, $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_k)$ is a vector of k unknown parameters measuring the influence of the k covariates $\mathbf{x}' = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$ and, t represents the univariate survival times of the units.

The ratio of the hazard function of two individuals, i and j, with $i \neq j$ where $i, j = 1, \ldots, n$, with of different covariates vector is given by

$$\begin{aligned}\tau(t|\mathbf{x}_i,\mathbf{x}_j) &= \frac{h(t|\mathbf{x}_i)}{h(t|\mathbf{x}_j)} = \frac{\lambda\exp(\alpha t+\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\alpha t+\mathbf{x}_i'\boldsymbol{\beta})}\frac{1+\exp(\alpha t+\mathbf{x}_j'\boldsymbol{\beta})}{\lambda\exp(\alpha t+\mathbf{x}_j'\boldsymbol{\beta})} \\ &= \frac{1+\exp(\alpha t+\mathbf{x}_j'\boldsymbol{\beta})}{1+\exp(\alpha t+\mathbf{x}_i'\boldsymbol{\beta})}\exp\big[(\mathbf{x}_i-\mathbf{x}_j)'\boldsymbol{\beta}\big].\end{aligned} \tag{2.2}$$

Note that the time effect does not disappear in (2.2) and hence the non-proportionality becomes evident.

From (2.1), the survival function is given by

$$S(t|\lambda,\alpha,\boldsymbol{\beta}) = \left\{\frac{1+\exp(\alpha t+\mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})}\right\}^{-\lambda/\alpha}. \tag{2.3}$$

The behavior of the hazard function (2.1) takes several forms, according to the value of α: for $\alpha > 0$, the hazard function is increasing; for $\alpha < 0$, the hazard function is decreasing; for $\alpha = 0$, the hazard function is constant. The survival function given in (2.3) also has its behavior determined by the value of α. For $\alpha > 0$, $S(0|\lambda,\alpha,\boldsymbol{\beta}) = 1$ and $S(\infty|\lambda,\alpha,\boldsymbol{\beta}) = \lim_{t\longrightarrow\infty} S(t|\lambda,\alpha,\boldsymbol{\beta}) = 0$, in other words, the survival function is proper, and for $\alpha < 0$, $S(0|\lambda,\alpha,\boldsymbol{\beta}) = 1$ and $S(\infty|\lambda,\alpha,\boldsymbol{\beta}) \neq 0$, so we have that the survival function is improper, for example, when $\alpha < 0$ we have a model for cure rate or long duraction, and the cure fraction, p, is given by

$$p = \big(1+\exp(\mathbf{x}'\boldsymbol{\beta})\big)^{\lambda/\alpha}.$$

The advantage of the GTDL model regarding the Cox model is that the ratio of the hazard function of two individuals changes over time and, the GTDL model can assume a behavior of a long duration model. Moreover, the GTDL arises from

the need for removing an inconvenient constraint on the hazard function imposed by the TDL (Time-Dependent Logistic) model given by $0 < h_{\mathrm{TDL}}(t|\alpha, \beta, x) \leq 1$, for all $t > 0$. This fact led to the incorporation of the parameter $\lambda > 0$ into the hazard function and the consequent removal of β_0 from the model.

The evolution of the TDL model to the GTDL model is the inclusion of the parameter λ in the TDL model with the objective that the hazard function is not limited in the interval $(0, 1)$. Mackenzie (2002) justifies the removal of the parameter β_0 due to the inclusion of λ, because the role of the parameters are interchangeable, so only one parameter is needed in the model. Louzada-Neto et al. (2010) presented a Bayesian approach for the GTDL model and Mackenzie (1997) used this model when the data set presents recurrent events.

2.2 GTDL frailty model

From the GTDL model given in the equation (2.1), the hazard function of the ith individual with the frailty term v_i multiplicative is given by

$$h_i(t|\alpha, \boldsymbol{\beta}, \lambda, v_i) = v_i \frac{\lambda \exp(\alpha t + \mathbf{x}_i'\boldsymbol{\beta})}{1 + \exp(\alpha t + \mathbf{x}_i'\boldsymbol{\beta})}, \tag{2.4}$$

interpreted as the conditional hazard function of the ith individual given v_i. The conditional survival function is given by

$$S_i(t|\alpha, \boldsymbol{\beta}, \lambda, v_i) = \left(\frac{1 + \exp(\alpha t + \mathbf{x}_i'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i'\boldsymbol{\beta})}\right)^{-\lambda v_i/\alpha}. \tag{2.5}$$

Note that, if we build the likelihood function using the hazard and survival functions given in (2.4) and (2.5), respectively, would have a likelihood function with more parameters to be estimated than observations. Moreover, we assume that the frailty v_i is a random variable independent and identically distributed with density function gamma$(1/\theta, 1/\theta)$ (see Wienke, 2011). This parametrization is considered to obtain $\mathrm{E}(v_i) = 1$ and $\mathrm{Var}(v_i) = \theta$. The unconditional survival function is given by

$$S(t) = \int_0^\infty S(t|\alpha, \boldsymbol{\beta}, \lambda, v_i) f(v_i)\, dv_i,$$

where $f(v_i)$ is the probability density function of the gamma. Giving a function $f(x)$, the Laplace transforming considers a function of real argument s is defined as

$$Q(s) = \int_0^\infty e^{-sx} f(x)\, dx. \tag{2.6}$$

The reason why the Laplace transform is very useful in this situation is because it has the same shape as the unconditional survival function. In the equation (2.6),

suppose $f(x)$ is the density function of the frailty variable V and s is the cumulated hazard function $H(t)$. Then we obtain,

$$S(t) = \int_0^\infty e^{-H(t)v} f(v)\, dv = Q\big(H(t)\big).$$

The Laplace transform of the gamma$(1/\theta, 1/\theta)$, considering s a real argument, is given by

$$Q(s) = \left(\frac{1/\theta}{1/\theta + s}\right)^{1/\theta} = (1+\theta s)^{-1/\theta}. \tag{2.7}$$

Substituting $s = H(t)$ in the equation (2.7), we obtain the unconditional survival function, given by

$$S(t|\alpha, \boldsymbol{\beta}, \lambda, \theta) = \left[1 + \frac{\lambda\theta}{\alpha} \log\left(\frac{1+\exp(\alpha t + \mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})}\right)\right]^{-1/\theta}, \tag{2.8}$$

and the correspondent hazard function is given by

$$h(t|\alpha, \boldsymbol{\beta}, \lambda, \theta) = \frac{\lambda \exp\big(\alpha t + \mathbf{x}'\boldsymbol{\beta}\big)}{[1 + \frac{\lambda\theta}{\alpha} \log(\frac{1+\exp(\alpha t+\mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})})](1+\exp(\alpha t + \mathbf{x}'\boldsymbol{\beta}))}. \tag{2.9}$$

The hazard function given in (2.9) takes unimodal form. The cure fraction, p, of the survival function given in (2.8), is given by

$$p = \left[1 + \frac{\lambda\theta}{\alpha} \ln\left(\frac{1}{1+\exp(\mathbf{x}'\boldsymbol{\beta})}\right)\right]^{-1/\alpha},$$

that is regulated by a regression model using covariates that may or may not be different from those used in the regression model of the hazard.

2.3 Inference

Let T be a random variable representing the failure time of the unit i. The likelihood function for the censored data is constructed from the equations (2.8) and (2.9) and is given by

$$\begin{aligned} L(\alpha, \boldsymbol{\beta}, \lambda, \theta|\text{dados}) &= \prod_{i=1}^{n} \big[h(t_i; \lambda, \alpha, \boldsymbol{\beta}, \theta)\big]^{\delta_i} \big[S(t_i; \lambda, \alpha, \boldsymbol{\beta}, \theta)\big] \\ &= \prod_{i=1}^{n} \left[\frac{\lambda \exp(\alpha t + \mathbf{x}'\boldsymbol{\beta})}{1+\exp(\alpha t + \mathbf{x}'\boldsymbol{\beta})}\right]^{\delta_i} \\ &\quad \times \prod_{i=1}^{n} \left[1 + \frac{\lambda\theta}{\alpha} \log\left(\frac{1+\exp(\alpha t + \mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})}\right)\right]^{-(1/\theta+\delta_i)}, \end{aligned} \tag{2.10}$$

where δ_i is a indicator variable of censoring, assuming 1 for observed failure and 0 for censoring.

Denoting $l(\alpha, \boldsymbol{\beta}, \lambda, \theta|\text{data}) = \log(L(\alpha, \boldsymbol{\beta}, \lambda, \theta|\text{data}))$, we obtain

$$\begin{aligned} l(\alpha, \boldsymbol{\beta}, \lambda, \theta|\text{data}) = \log(\lambda) \sum_{i=1}^{n} \delta_i + \sum_{i=1}^{n} \delta_i (\mathbf{x}'\boldsymbol{\beta} + \alpha t_i) \\ - \sum_{i=1}^{n} \delta_i \log(1 + \exp(\alpha t + \mathbf{x}'\boldsymbol{\beta})) \\ - \sum_{i=1}^{n} (\delta_i + 1/\theta) \left\{ \log\left[1 + \frac{\theta\lambda}{\alpha} \log\left(\frac{1 + \exp(\alpha t + \mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})} \right) \right] \right\}. \end{aligned} \tag{2.11}$$

The maximum likelihood estimates (MLEs) are obtained by direct maximization of equation (2.10). The asymptotic confidence intervals are obtained by considering the maximum likelihood estimates and the inverse of the observed information matrix.

Model comparison between GTDL and GTDL frailty models is made by considering the Akaike information criterion (AIC) (Akaike, 1974) and Bayesian information criterion (BIC) (Schwarz, 1978). For the GTDL frailty model the AIC and BIC criteria are given, respectively, by $-2l(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\lambda}, \widehat{\theta}|\text{dados}) + 2q$ and $-2l(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\lambda}, \widehat{\theta}|\text{dados}) + q\log(n)$, where $(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\lambda}, \widehat{\theta})$ denotes the MLE, q is the number of parameters in the model and n is the size sample.

3 Simulation study

The simulation study main concern is to assess the bias and mean squared error (MSE) of the MLEs as well as the coverage probabilities of the asymptotic confidence intervals for the parameters of the GTDL frailty model.

We generated 1000 samples for each sample size ($n = 50, 100, 300$ and 500). The parameters were fixed at $\alpha = 0.10$, $\beta = -3.00$, $\lambda = 0.50$ and $\theta = 0.50$. A dummy covariate was generate from a Bernoulli distribution with success probability equal to 0.50. The censored times were generated from an exponential distribution with parameter equals to 110 for 10% of censoring and equal to 27 for 30% of censoring. For each sample, we obtain the MLEs, calculate their bias and MSEs, and the asymptotic 95% confidence intervals. We check if the value of the true parameter was contained in the confidence interval and withheld the number of times this occurs. Thus, the coverage probability was found as the quotient between the number of intervals containing the true parameter value and the total number of intervals constructed, here equals to 1000.

The bias and MSEs are presented in the Table 1. The bias and MSEs decrease tending to zero when the sample size increases.

Figure 1 shows the empirical coverage probabilities according to each sample size and the different censoring percentages. In the figure, the two black horizontal

Table 1 *Bias and MSE of the MLEs for* 0% / 10% / 30% *of censoring*

		Bias				MSE			
Size	Censoring	α	β	θ	λ	α	β	θ	λ
$n = 50$	0	−0.007	0.048	1.57	0.051	0.005	0.495	13.444	0.129
	10	−0.016	0.055	1.731	0.066	0.011	0.625	15.882	0.133
	30	−0.047	0.338	2.218	0.046	0.401	1.524	21.440	0.143
$n = 100$	0	−0.006	0.058	0.396	0.029	0.002	0.223	2.350	0.061
	10	−0.004	0.032	0.650	0.029	0.002	0.261	4.414	0.066
	30	−0.008	0.050	0.850	0.035	0.003	0.308	6.500	0.070
$n = 300$	0	−0.001	0.013	0.064	0.004	0.000	0.070	0.064	0.018
	10	−0.001	0.003	0.080	0.015	0.000	0.079	0.090	0.018
	30	−0.002	0.020	0.109	0.006	0.000	0.101	0.245	0.023
$n = 500$	0	0.000	−0.002	0.027	0.006	0.000	0.046	0.032	0.012
	10	0.000	0.005	0.059	0.007	0.000	0.046	0.05	0.011
	30	0.000	0.006	0.071	0.006	0.000	0.053	0.084	0.014

lines represent the lower and upper limits of 95% confidence interval of the nominal coverage probability. The empirical coverages are closer to the nominal ones for increasing sample sizes, but the empirical coverage probability for θ is higher than expected.

Now we turn our interest to discover the behavior of the GTDL model and the GTDL model with fragility when an important covariate is ignored in the estimation procedure by performing a simulation study with different sample sizes. Lifetimes were generated from the GTDL model assuming $\alpha = 0.1$, $\beta_1 = -3$, $\beta_2 = -1$ and $\lambda = 0.5$, with two covariates were considered, X_1 was generated from a binomial distribution with $p = 0.5$ and X_2 was generated from a standard normal distribution. We then fitted both models using only one covariate at a time. The idea is to verify if the models are be able to capture information from an important covariate with is purposely left out of the model fit. The procedure was repeated 1000 times for each set up. The mean of the MLEs are presented in Table 2. Clearly the estimates obtained via the GTDL model with fragility are closer to the true parameter values. We were expecting such result since the GTDL model with frailty aims to capture information from important covariate that were not observed.

4 Lung cancer data

In this section, we illustrated our model applied to a real data set. The dataset consisting of a population-based prospective study of incident cases of lung cancer diagnosed in Northern Ireland in one year (Wilkinson, 1995). The study was conducted between 01/10/1991 and 30/09/1992. For the analysis, we included only

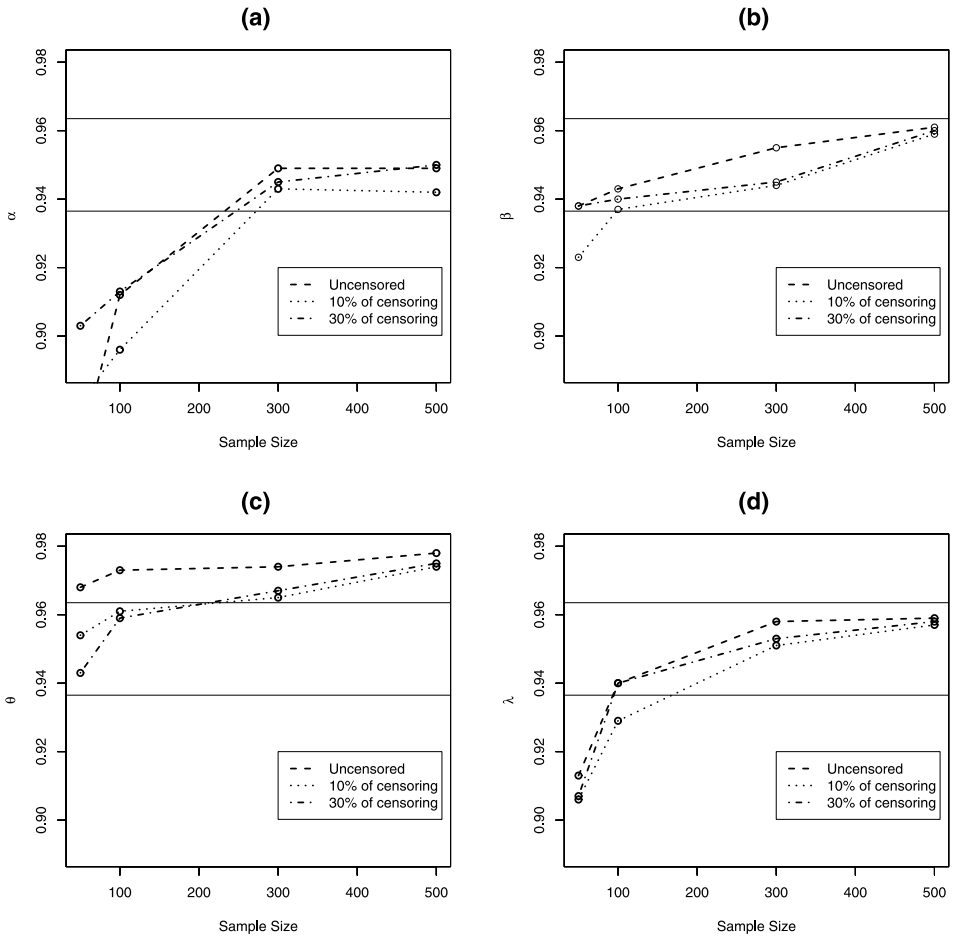

Figure 1 *Probability of coverage of the* 95% *confidence intervals for* α, β, θ *and* λ.

Table 2 *Mean of the MLEs of the GTDL with frailty/GTDL*

n	θ	α	β	λ
		with X_1		
50	0.095/– – –	0.061/0.077	−2.514/−2.692	0.466/0.519
100	0.095/– – –	0.054/0.068	−2.470/−2.649	0.468/0.513
300	0.073/– – –	0.054/0.064	−2.484/−2.613	0.461/0.493
500	0.077/– – –	0.051/0.061	−2.470/−2.607	0.457/0.489
		with X_2		
50	0.056/– – –	0.046/0.051	−0.967/−1.004	0.221/0.232
100	0.044/– – –	0.034/0.038	−0.951/−0.997	0.222/0.231
300	0.013/– – –	0.030/0.032	−0.912/−0.925	0.212/0.217
500	0.077/– – –	0.030/0.033	−0.950/−0.958	0.214/0.234

Table 3 *Information of the covariates*

Covariate	Type	Representation
Age (in years)	Continuous	β_1
Sex	Binary	β_2
Treatment group	Categorical	β_3, β_4, β_5 and β_6
WHO status	Categorical	β_7, β_8, β_9 and β_{10}
Cell type	Categorical	β_{11}, β_{12} and β_{13}
Sodium level	Categorical	β_{14}
Albumen level	Categorical	β_{15}
Metastases	Categorical	β_{16} and β_{17}
Smoking	Categorical	β_{18} and β_{19}

individuals who had information on all covariates, and therefore, we analyzed the lifetimes of 751 patients (in months). The observed covariates, type and representation are presented in Table 3 according to the dummying structuring.

The cumulative log-hazard plot in Figures 2 indicates non-proportionality for the covariates: treatment, WHO status, cell type, metastasis and smoking, where $H(t)$ was estimated by using the usual Nelson–Aalen–Breslow estimator.

Besides the covariates in Table 3, other factors related to lung cancer lifetime are certain chemical agents, dietary factors, chronic obstructive pulmonary disease, genetic factors and family history of lung cancer (Instituto Nacional do Câncer José Alencar Gomes da Silva, 2014). All these factors were not measured, motivating for the use of the proposed GTDL frailty model in order to capture the influence of such factors. Table 4 presents the MLEs, their standard deviation (SD) and 95% confidence intervals (CI) based on the fitting of the GTDL model and the GTDL frailty model. The natural consequence of not considering the presence of the frailty is the underestimation of the SD and consequently the underestimation of the correspondent CI width. Note that the parameter that measure the heterogeneity (θ) is significant. Further note that the covariate parameters β_4 and β_5 related to the type of treatment, β_{11} and β_{13} related to the cell type, β_{16} related to the presence of metastases, and β_{18} and β_{19} related to the use of tobacco are not significant.

Table 3 also shows the AIC and BIC model comparison criteria for the two fitted models together with the value of the log-likelihood values calculated on the MSEs. Both criteria indicate evidence in favor to the GTDL frailty model.

5 Final comments

In this paper, we consider an extension of GTDL model, the GTDL frailty model, finding its hazard, survival and density functions. The simulation study shows that the MLEs are unbiased, the MSEs decrease with the increasing of the sample size

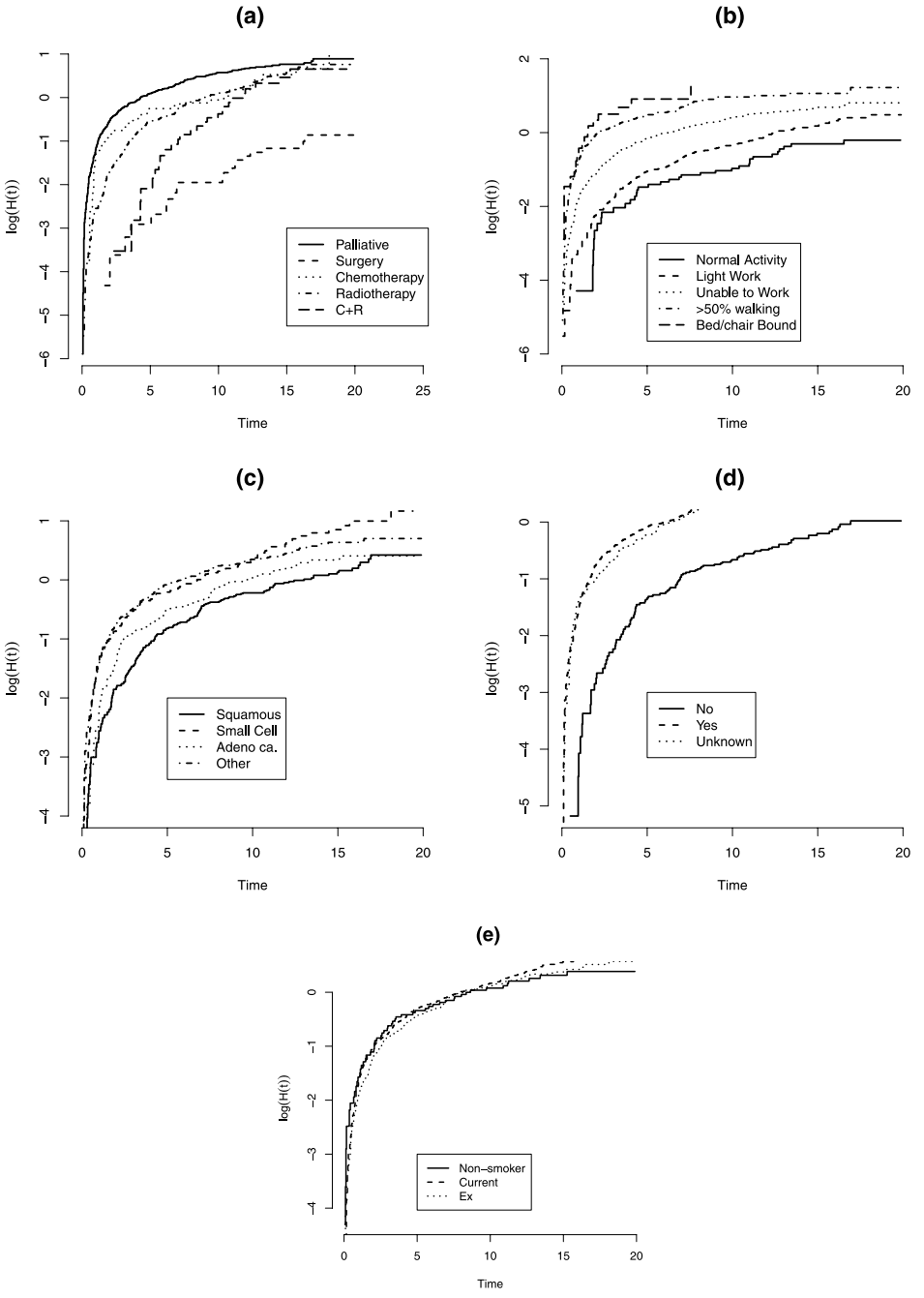

Figure 2 *Cumulative log-hazard plots for the covariates*: *treatment in* (a), *WHO status in* (b), *cell type in* (c), *metastases in* (d) *and smoking in* (e).

Table 4 *Results of fitting the GTDL frailty and GTDL models*

	GTDL Frailty Model			GTDL Model		
Parameters	MLE	SD	CI (95%)	MLE	SD	CI (95%)
θ	0.598	0.128	(0.347; 0.849)			
α	0.139	0.029	(0.082; 0.195)	0.016	0.013	(−0.011; 0.043)
β_3	1.718	0.452	(0.833; 2.604)	1.087	0.327	(0.446; 1.728)
β_4	−0.145	0.499	(−1.122; 0.832)	−0.172	0.385	(−0.927; 0.583)
β_5	0.756	0.428	(−0.083; 1.595)	0.328	0.301	(−0.261; 0.918)
β_6	1.185	0.441	(0.320; 2.049)	0.781	0.320	(0.153; 1.409)
β_7	−4.119	0.566	(−5.228; −3.010)	−3.142	0.373	(−3.873; −2.412)
β_8	−3.979	0.537	(−5.032; −2.927)	−2.984	0.340	(−3.650; −2.318)
β_9	−3.207	0.517	(−4.221; −2.193)	−2.517	0.334	(−3.170; −1.864)
β_{10}	−2.117	0.571	(−3.236; −0.997)	−1.790	0.357	(−2.489; −1.091)
β_{11}	−0.180	0.187	(−0.546; 0.185)	−0.233	0.131	(−0.489; 0.023)
β_{12}	1.094	0.345	(0.418; 1.770)	0.797	0.248	(0.311; 1.282)
β_{13}	0.024	0.226	(−0.419; 0.467)	0.129	0.169	(−0.202; 0.461)
β_{14}	−0.463	0.154	(−0.764; −0.161)	−0.366	0.110	(−0.581; −0.151)
β_{15}	−0.767	0.167	(−1.094; −0.442)	−0.524	0.115	(−0.749; −0.299)
β_{16}	−0.311	0.231	(−0.7626; 0.141)	−0.289	0.164	(−0.610; 0.032)
β_{17}	0.638	0.191	(0.264; 1.012)	0.520	0.136	(0.253; 0.787)
β_{18}	−0.137	0.266	(−0.659; 0.384)	−0.236	0.187	(−0.602; 0.131)
β_{19}	0.079	0.154	(−0.223; 0.381)	0.132	0.110	(−0.083; 0.347)
λ	1.413	0.392	(0.645; 2.181)	1.082	0.233	(0.6256; 01.538)
log like	−1596.52			−1613.19		
AIC	3233.04			3264.39		
BIC	3325.49			3390.22		

and that for samples of reasonable size the coverage probabilities are close to the nominal. The application on the real data shows that the GTDL frailty model fitted better than the model without the frailty term. The fact that θ is significant indicates that not factors observed have influence on the lifetimes.

We can conclude that the use of the model without frailty can lead to wrong interpretations. The parameter α, which measure the effect of the time in the GTDL frailty model, is significant and positive, then we have the effect of increasing the individual hazard function, anticipating the occurrence of failure, which does not happen in the model without frailty.

Although model comparison between GTDL and GTDL frailty models made by considering the AIC and the BIC, a hypothesis test could be useful to decide between the models. However, the asymptotic distribution of the test statistic should be carefully addressed since the hypothesis in this case is on the boundary of the parameter space. This matter should be investigated further (see Wienke, 2011).

In this paper, GTDL models with frailty are used to capture information from important covariates that are missing, however, at least in principle, it can be

adapted for recurrent event data in order to describe the dependence among the individual lifetimes, as well as for lifetime data in presence of long-term survivals. Such extension is however not within the scope of this paper.

Acknowledgments

The authors are grateful to Gilbert MacKenzie (University of Limerick, Ireland) for providing the Northern Ireland lung cancer data. The research of Eder A. Milani was supported by FAPESP-Brazil. The research of Francisco Louzada is supported by CNPq-Brazil.

References

Aalen, O. and Tretli, S. (1999). Analysing incidence of testis cancer by means of a frailty model. *Cancer Causes and Control* **10**, 285–292.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–723. MR0423716

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151. MR0501698

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Ser. B* **34**, 187–220. MR0341758

Etezadi-Amoli, J. and Ciampi, A. (1987). Extended hazard regression for censored survival data with covariates: A spline approximation for the baseline hazard function. *Biometrics* **43**, 181–192.

Henderson, R. and Oman, P. (1999). Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society, Ser. B* **61**, 367–379. MR1680322

Ibrahim, J., Chen, M. and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer. MR1876598

Instituto Nacional do Câncer José Alencar Gomes da Silva (2014). Prevenção câncer de pulmão. Available at http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/pulmao/prevencao.html. Accessed on July 30, 2014.

Louzada-Neto, F. (1997). Extended hazard regression model for reliability and survival analysis. *Lifetime Data Analysis* **3**, 367–381.

Louzada-Neto, F. (1999). Polyhazard models for lifetime data. *Biometrics* **55**, 1281–1285.

Louzada-Neto, F., Cremasco, C. P. and Mackenzie, G. (2010). Sampling-based inference for the generalized time-dependent logistic hazard model. *Journal of Statistical Theory and Applications* **9**, 169–184.

Mackenzie, G. (1996). Regression models for survival data: The generalized time-dependent logistic family. *The Statistician* **45**, 21–34.

Mackenzie, G. (1997). On a non-proportional hazards regression model for repeated medical random counts. *Statistics in Medicine* **16**, 1831–1843.

Mackenzie, G. (2002). A logistic regression model for survival data. In *17th International Workshop in Statistical Modelling* 443–450. Chania, Crete: Statistical Modelling in Society.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Ser. B* **44**, 414–422. MR0693242

Prentice, R. L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–179. MR0497517

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464. MR0468014

Tomazella, V. L. D., Rodrigues, J., Martins, C. B. and Bernardo, J. M. (2008). Inference on the univariate frailty model: A Bayesian reference analysis approach. *AIP Conference Proceedings* **1073**, 340–347.

Wienke, A. (2011). *Frailty Models in Survival Analysis*. Boca Raton, FL: Chapman & Hall/CRC Biostatistics Series. MR2682965

Wilkinson, P. (1995). Lung cancer in Northern Ireland. MSc thesis, Queen's Univ. Belfast, Belfast.

E. A. Milani
V. L. D. Tomazella
T. C. M. Dias
Universidade Federal de São Carlos
Rodovia Washington Luís
km 235 - SP-310
São Carlos - São Paulo
CEP 13565-905
Brasil
E-mail: edinhomilani@hotmail.com
vera@ufscar.br
dtmd@ufscar.br

F. Louzada
Universidade de São Paulo
São Carlos-SP
Brazil
E-mail: louzada@icmc.usp.br
URL: http://www.icmc.usp.br