SURVIVAL PARAMETRIC MODELS FOR MISMEASURED OUTCOMES

Magda C. Pires[1,3], Enrico A. Colosimo[1,3], Arlaine A. Silva[2]

Resumo: *A análise de dados de sobrevivência consiste no estudo do tempo até a ocorrência de um evento de interesse como, por exemplo, a morte de um paciente, cura ou recidiva de uma doença. Quando o tempo exato de ocorrência não é conhecido, mas sabe-se que ele aconteceu no intervalo entre duas avaliações consecutivas do indivíduo, estamos diante de um estudo com censura intervalar. A detecção do evento depende da qualidade dos testes aplicados: um indivíduo pode ser diagnosticado como doente quando na verdade ele está sadio ou um indivíduo doente pode ser diagnosticado como sadio. Nesses casos, ao utilizar métodos tradicionais de Análise de Sobrevivência, estimativas viciadas para os parâmetros da distribuição do tempo de falha são obtidas [Paggiaro e Torelli (2004)]. Apresentamos, então, uma proposta que incorpora a sensibilidade e a especificidade do teste a modelos probabilísticos de Análise de Sobrevivência com dados grupados (caso especial de censura intervalar em que todas as unidades são avaliadas nos mesmos instantes). Avaliamos o caso especial do Modelo Weibull de Riscos Proporcionais e os estudos de simulação Monte Carlo demonstraram que, quando a sensibilidade e a especificidade do teste são conhecidas, o método proposto é bastante eficiente, pois suas estimativas apresentam menor vício relativo do que aquelas fornecidas pelo método tradicional. Apresentamos ainda uma aplicação do modelo proposto a dados de tempo de vida de mangueiras.*

Palavras-chave: *Análise de Sobrevivência, Modelos Paramétricos, Erros de Classificação, Sensibilidade, Especificidade.*

1 Introduction

In some survival studies it is common to observe the combination of interval censoring and information bias. Interval censoring occurs when the event of interest is not observed exactly but it is known to occur within some time interval. A usual situation of interval censoring happens when patients in a clinical trial or longitudinal study have periodic follow-up and the event is observed throughout a diagnostic test. Many diagnostic tests are not perfectly sensitive and specific. That is, they may not indicate the true event status. This fact is known in epidemiology as information bias. Example of imperfect diagnostic include is Elisa test for HIV.

Cox regression analysis considering mismeasured outcomes has been treated in the literature. Snapinn (1998), for example, considered auxiliary variables and Meier et. al (2003) extended the discrete proportional hazards of Kalbfleisch and Prentice (2002) to incorporate misclassified responses.

We extend parametric survival models to include diagnostic errors in the event of interest. Parametric models are used in a day-by-day basis in areas, such as engineering and agronomy, where noise variables can be controlled by using experimental design. Many new parametric models have been proposed lately in literature showing the importance of them in the survival data analysis [see, for example, Cordeiro and Castro, 2011)].

[1]DEST-Universidade Federal de Minas Gerais, magda@est.ufmg.br, enricoc@est.ufmg.br

[2]arlaineas@gmail.com

[3]Agradecimento à FAPEMIG e à PRPq/UFMG (Programa Institucional de Auxílio à Pesquisa de Doutores Recém-Contratados) pelo apoio financeiro.

2 Methodology

Suppose that subjects enter a study at time 0 and during a follow-up period they are regularly screened at times $t_1, \ldots, t_k$ to detect the occurrence of a disease or a condition of interest (grouped survival data). When the event is detected, the subject is excluded from the study. If the used diagnostic test is not perfectly sensitive and specific, the outcome may be measured with error. The test sensitivity θ denotes the probability that the test detects the event when in fact it did occurred, and the specificity ϕ is probability that the test do not detect the event when in fact it did not occurred. In this context, $(t_{j-1}, t_j]$ is the observed failure interval time for the i-th subject with probability

$$\phi^{(j-1)}(1-\phi)S(t_j|\mathbf{x}_i) + \theta \sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{(j-l)}[S(t_{l-1}|\ \mathbf{x}_i) - S(t_l|\ \mathbf{x}_i)],$$

and $(t_{j-1}, t_j]$ is the observed censoring interval time for the i-th subject with probability

$$\phi^{j} S(t_j|\mathbf{x}_i) + (1-\theta) \sum_{l=1}^{j} \phi^{l-1}(1-\theta)^{(j-l)}[S(t_{l-1}|\mathbf{x}_i) - S(t_l|\ \mathbf{x}_i)].$$

where $S(t_j|\mathbf{x}_i)$ is the survival function at time t_j.

Then, the likelihood function for grouped data subject to misclassification is given by

$$L = \prod_{j=1}^{k} \prod_{r \in R_j} \left(\phi^{(j-1)}(1-\phi)S(t_j|\mathbf{x}_r) + \theta \sum_{l=1}^{j} \phi^{(l-1)}(1-\theta)^{(j-l)}(S(t_{l-1}|\mathbf{x}_r) - S(t_l|\mathbf{x}_r)) \right)^{n_{j,r}} \times$$
$$\times \left(\phi^{j} S(t_j|\mathbf{x}_r) + (1-\theta) \sum_{l=1}^{j} (S(t_{l-1}|\mathbf{x}_r) - S(t_l|\mathbf{x}_r)) \phi^{(l-1)}(1-\theta)^{(j-l)} \right)^{m_{j,r}}, \quad (1)$$

where R_j is the subset of those subjects with $\mathbf{x}_r$ covariates and failure or censoring interval time equal $(t_{j-1}, t_j]$, and $n_{j,r}$ $(m_{j,r})$ is the number of observed failure (censoring) in that interval time among such subjects.

2.1 Weibull Proportional Hazards Model for Mismeasured Outcomes

Likelihood function (1) holds for any parametric model with survival function $S(\cdot)$. For the Weibull proportional hazards model, where $S(t_j|\mathbf{x}_i) = \exp\left\{-\lambda \exp(\beta' \mathbf{x}_i)\, t_j^{\gamma}\right\}$ [Collett (2003)], the likelihood function (1) turns to be

$$L = \prod_{j=1}^{k} \prod_{r \in S_j} \left(\phi^{(j-1)}(1-\phi) b_{j,r} + \theta a_{j,r} \right)^{n_{j,r}} \left(\phi^{j} b_{j,r} + (1-\theta) a_{j,r} \right)^{m_{j,r}}, \quad (2)$$

where $a_{j,r} = \sum_{l=1}^{j} (b_{l-1,r} - b_{l,r})\, \phi^{l-1}\, (1-\theta)^{j-l}$ and $b_{j,r} = \exp\left\{-\lambda \exp(\beta' \mathbf{x}_r)\, t_j^{\gamma}\right\}$.

Maximum likelihood estimates for β are obtained by numerically maximizing the likelihood function (2) and its asymptotic variance is given by $I(\beta)^{-1}$ where $I(\beta)$ is the observed information matrix (not shown).

3 Monte Carlo Simulations

In order to assess the relative performance of the proposed methodology (Prop.) compared to the standard one that ignores measurement errors (Stan.), data sets were generated from a

Weibull ($\lambda = 0.25$, $\gamma = 2$) distribution with a single binary covariate such that $\beta_1 = 1.5$. Subjects were screened at times $0.5, 1.0, 1.6, 2.0, 2.7, 3.0$ by a imperfect diagnostic test (sensitivity θ and specificity ϕ were 0.8, 0.85 and 0.95). Approximately 10% of subjects were censored before $t_6 = 3.0$ and 10% at the end of follow-up period. $S = 1000$ simulations were performed for each scenario and samples of 1200 observations per simulation were used.

Both models were evaluated with respect to percent bias, standard error and mean square error. The last two measures are presented just for the regression coefficient β_1. A summary of the obtained results is presented in Table 1.

Tabela 1: Summary of simulation results

		% Bias γ		% Bias λ		% Bias β_1		Std.Err.		MSE	
ϕ	θ	Prop.	Stan.	Prop.	Stan.	Prop.	Stan.	Prop.	Stan.	Prop.	Stan.
0,95	0,95	3,857	-6,515	-21,489	12,566	4,131	-15,877	0,055	0,043	0,007	0,059
0,95	0,90	3,897	-6,612	-21,239	10,156	4,196	-18,277	0,063	0,047	0,008	0,077
0,95	0,80	4,226	-7,025	-21,076	4,815	4,536	-23,616	0,080	0,052	0,011	0,128
0,90	0,95	4,119	-14,154	-21,905	53,205	4,732	-29,263	0,081	0,052	0,012	0,195
0,90	0,90	4,632	-14,125	-22,534	49,379	5,546	-31,062	0,084	0,051	0,014	0,220
0,90	0,80	4,536	-15,169	-21,550	44,461	5,274	-36,603	0,106	0,056	0,017	0,305
0,80	0,95	5,124	-24,376	-22,902	146,303	6,060	-48,659	0,121	0,057	0,023	0,536
0,80	0,90	5,447	-24,796	-23,141	142,574	6,517	-50,774	0,136	0,056	0,028	0,583
0,80	0,80	5,904	-25,996	-22,475	136,192	6,527	-55,811	0,158	0,055	0,034	0,704

Results corroborates the fact that ignoring error probabilities produces biased estimates, since absolute percent bias under the standard model is greater than those produced by the model for mismeasured outcome. However, standard error of the estimatives is greater in this model. We also note that, under the proposed model, shape parameter of the Weibull distribution and the regression coefficient were overestimated (positive percent bias) while scale parameter was underestimated (negative percent bias). The opposite occurs under the standard model. Besides that, under both models, bias and standard error increase as the quality of the test, specially specificities, decreases.

4 Numeric Example

An experiment was conducted in a completely randomized block design with five blocks and six treatments in a 6 x 7 factorial design involving six different scions grafted (Extrema, Oliveira, Pahiri, Imperial, Carlota and Bourbon) on seven different stocks (Espada, Extrema, Oliveira, Carlota, Bourbon, Coco and Pahiri), totaling 210 experimental units. The aim of the experiment was to determine the scion-stock combination most resistant to a disease (*seca* of the mango tree) caused by the *Ceratocystis fimbriata* fungus. The experimental study began in 1971; the site of the experiment was visited 12 times in the years 1973, 1974, 1975, 1981, 1983, 1985-1990 and 1992 and the condition of each experimental unit (alive or dead) was registered. A more detailed description of this data set can be found in Colosimo et. al (2000).

In this work the interest focuses only on lifetimes for the six scions. However, the lifetimes are not exactly known. Since data are available for all units in every visit and many mango trees die in the same time interval, we are deling with grouped survival data.

Additionally, suppose that detection of a dead mango tree may be subject to a known error rate, in the sense that a dead mango tree may be erroneously considered alive (the opposite is unlikely). So, the method proposed in section 3.1 is fitted using some values of sensitivity between 0.5 and 1.0. Results are presented in Table 2.

We realized that the standard error increases as the sensitivity decreases, but estimates of risk suffer little variation. Hence, the estimate significance generally decreases (see, for example, β_5). The hazard ratio estimates obtained ($\exp(\hat{\beta})$) can be interpreted as follows: the risk of Extrema scions mango tree die is 1.53 $\left[(\exp(\hat{\beta}_1))^{-1}\right]$ times the risk of the Oliveira scions mango

Tabela 2: Estimates for mango tree survival data

		β_1: Oliveira	β_2: Pahiri	β_3: Imperial	β_4: Carlota	β_5: Bourbon
θ=1.0, ϕ=1	$\hat{\beta}$	-0,423	0,044	-0,211	-0,264	0,504
	$se(\hat{\beta})$	0,128	0,063	0,100	0,103	0,169
	$\exp(\hat{\beta})$	0,655	1,044	0,809	0,768	1,655
θ=0.9, ϕ=1	$\hat{\beta}$	-0,422	0,045	-0,211	-0,262	0,506
	$se(\hat{\beta})$	0,128	0,063	0,102	0,103	0,203
	$\exp(\hat{\beta})$	0,656	1,046	0,810	0,770	1,659
θ=0.7, ϕ=1	$\hat{\beta}$	-0,425	0,047	-0,210	-0,254	0,525
	$se(\hat{\beta})$	0,131	0,065	0,105	0,104	0,243
	$\exp(\hat{\beta})$	0,654	1,048	0,811	0,776	1,690
θ=0.5, ϕ=1	$\hat{\beta}$	-0,434	0,048	-0,214	-0,256	0,567
	$se(\hat{\beta})$	0,132	0,066	0,106	0,105	0,262
	$\exp(\hat{\beta})$	0,648	1,049	0,807	0,774	1,763

tree, the risk of Pahiri scions mango tree die is 1.04 $\left[(\exp(\hat{\beta}_2))\right]$ times the risk of the Extrema scions mango tree, and so on.

5 Final Remarks

In this paper, we presented a parametric model that incorporates classification errors into survival analysis, which can be used with a great variety of probability distributions. We analyzed the Weibull distribution special case. Monte Carlo simulations demonstrated that the parametric model for mismeasured outcome produces small percent bias and greater standard error relative to the model that assumes perfect sensitivity and specificity. Besides that, lower specificity induces greater bias in estimation than lower sensibility (similar results were obtained by Meier et. al (2003)) and the higher the censoring percentage, higher the percent bias (results not shown).

The model presented in this paper assumes that the sensitivity and specificity of the diagnostic test are known and we also demonstrated that it is important to have good estimates for these parameters, since the performance of the method is reduced when one of them (specially specificity) is misspecified (results not shown).

Referências

[1] Paggiaro, A., Torelli, N., The effect of classification errors in survival data analysis, *Statistical Methods and Applications*, **13**, 213-225, 2004.

[2] Snapinn, S. M., Survival analysis with uncertain endpoints, *Biometrics*, **54**, 209-218, 1998.

[3] Meier, A. S., Richardson, B. A., Hughes, J. P., Discrete proportional hazards models for mismeasured outcomes, *Biometrics*, **59**, 947-954, 2003.

[4] Kalbfleisch, J. D., Prentice, R. L., *The statistical analysis of failure time data*, John Wiley, 2002.

[5] Cordeiro, G. M., Castro, M. A new family of generalized distributions, *Journal of Statistical Computation and Simulation*, **81(7)**, 883-898, 2011.

[6] Collett, D., *Modelling survival data in medical research*, Chapman-Hall, 2003.

[7] Colossimo, E. A., Chalita, L. V. A. S., Demétrio, C. G. B., Tests of proportional hazards and proportional odds models for grouped survival data, *Biometrics* **56**, 1233-1240, 2000.