Modelling the recurrence of bladder cancer

Gregorio Rubio[1], Cristina Santamaría[1], Belén García[1], and José Luis Pontones[2]

[1] Matemática multidisciplinar
Universidad Politécnica
Valencia (Spain)
(e-mail: crisanna@imm.upv.es, magarmo5@imm.upv.es, grubio@imm.upv.es)
[2] Hospital Universitario La Fe
Valencia (Spain)
(e-mail: pontones_jos@gva.es)

Abstract. The aim of this paper is to evaluate the risk of tumor recurrence of bladder cancer after surgical operation (TUR: Trans-urethral Resection). The prognostic significance of some clinical features in 454 patients with primary superficial bladder carcinoma is studied. The modelling procedure is featured within interval censored and right censored framework.
Keywords: bladder carcinoma, prognostic factors, recurrence, interval–censored survival data, generalized non-linear model, Cox model.

1 Introduction

Transitional bladder cancer represents about 2% of all human tumors. It supposes an important public health problem because it is biologically very aggressive and causes more than 130.000 deaths by year all around the world. Superficial bladder tumors are characterized by *recurrence* (reappearance of a new tumor) in 50-70% of cases. Although most recurrences are still superficial, *progression* to muscle invasive disease occurs in 10-30% of patients. Therefore, when superficial bladder tumor is diagnosed, it is important to identify patients who are at risk of disease recurrence and progression. If it were possible to define exactly which subset of superficial bladder tumors have more risk to recur and to progress, preemptive therapy could be used. Identifying the prognostic factors that determine that risk in each patient remains a subject of extensive research [Jaemal *et al.*, 2003], [Black *et al.*, 2002] and [Royston *et al.*, 2002].

Biotechnological advances have allowed us to use different therapeutic procedures (surgery, radiotherapy, chemotherapy, immunotherapy) successfully but still many patients suffer an unfavorable outcome without control of disease.

Multiple clinical and pathological variables are important in predicting outcome in patients with transitional bladder cancer, among which pathological stage and grade of differentiation are recognized as the most important [Zieger *et al.*, 1998], [Kurth *et al.*, 1995]. Therefore, an ideal prediction model

should combine stage, and grade, along with any other features shown to be associated with outcome in a multivariate model (histological characteristics, size, number of tumors, etc).

The TNM system (classification of 1997) is generally used to establish the stage of the bladder tumors [Hermanek and Sobin, 1998]:

Tis : tumor is limited to the mucosa and is flat (a carcinoma in situ).
Ta : tumor is papillary and it is limited to the mucosa.
T1 : tumor penetrates the lamina propia but not the muscle layer.
T2-T4 : tumor invades muscle and is staged from T2 to T4 according to the depth of infiltration of muscle tissue or the extent to which the surrounding tissue is affected.

Superficial bladder tumors (stages Ta and T1) have trend to produce recurrences (generally with similar stage). Tumors that invade the bladder muscle are highly aggressive and have a strong potential metastasize preferentially to regional lymph nodes, lungs, liver, and bone.

The *histologic grade* establishes according to the WHO (World Health Organization) 1999 classification [Hermanek and Sobin, 1998]:
G1: Urothelial carcinoma grade I (differentiated)
G2: Urothelial carcinoma grade II (intermediate differentiation)
G3: Urothelial carcinoma grade III (poor differentiated)

Well differentiated tumors (G1 grade) have generally low agressivity while poor differentiated tumors (G3 grade) are highly aggressive (cause many recurrences) [Millan *et al.*, 2000].

Prediction models can be used to counsel patients, determine the need for adjuvant therapy, stratify patients in risk groups, and develop appropriate postoperative surveillance programs tailored to risk for cancer progression. There are quite a few models in the medical literature, see [Millan *et al.*, 2000] for a little account. Nevertheless, many studies are based only on univariate analysis. Even if multivariate analysis is performed, usually the event of interest, for instance tumor recurrence, is recorded at scheduled screening times. It may be more convenient to consider arbitrarily interval-censored survival data because the exact time of the event of interest is not known. Our aim is to construct a prognostic model for predicting the outcome of superficial bladder cancer of transitional cells, within this framework. Then we perform the usual Cox model approach in order to compare.

In our study the time origin concern to the so called TUR (trans-urethral resection): a surgical endoscopic technique used to remove the macroscopic tumor from the inner of the bladder. The end-point is the first tumor recurrence.

The paper is organized as follows: in Section 2 the data on the survival times of 454 patients and their characteristics (explanatory variables) are described. In Section 3 we give a brief description of a method for analyzing interval–censored data proposed by Farrington [Farrington, 1996], and we

apply the method to our data base. In Section 4, a multivariate analysis is performed by using the *Cox proportional hazards* model.

We have used the packages S-PLUS ([Venables and Ripley, 2002]), SPSS and SAS ([Delwiche and Slaughter, 1998]).

2 Data and selection of variables

In this research, 454 patients from *La Fe* University Hospital from Valencia (Spain) were examined. They had primary superficial transitional cell carcinoma of the bladder initially treated with transurethral resection (TUR). The variable of interest was time (in days) from TUR to the first appearance of recurrence. The exact time of the recurrence will be unknown and the only information available concerns whether or not recurrence is identified when a patient visits the clinic. So, each individual may have a different time interval in which the recurrence has occurred and data are referred to as arbitrarily *interval-censored data*.

The period goes from 1973 to 2003. Variables considered for this study were: sex, age, tumor stage (pTa and pT1), tumor grade (G1, G2 and G3), number of tumors (one or more than one), tumor size (≤ 3 cm or > 3 cm) and treatment (Thiotepa, Adriamicine, Cisplatine, BCG and others treatments), see Table 1

3 Interval–censored analysis

The method for analyzing such data, assuming proportional hazards, is based on a non-linear model for binary data. The model is known as a *generalized non-linear model*[Farrington, 1996], see [Collett, 2003]:

The likelihood function for n observations may be expressed as:

$$\prod_{i=1}^{n+c} p_i^{y_i}(1-p_i)^{1-y_i} \tag{1}$$

where $y_1, y_2, \ldots, y_{n+c}$ are observations from a Bernoulli distribution with response probability p_i, $i = 1, 2, \ldots, n+c$, where c is the number of confined observations.

The survivor function is given by:

$$S_i(t) = S_0(t)^{exp(\beta' x_i)} \tag{2}$$

where $S_0(t)$ is the baseline survivor function and x_i is the vector of values of p explanatory variables for the ith individual, $i = 1, 2, \ldots, n$. The baseline survivor function will be modelled as a step function, where the steps occur at the k ordered censoring times, $t_1, t_2, \ldots, t_k$, where $t_1 < t_2 < \ldots < t_k$ (subset of times at which observations are interval–censored).

Variable	N patients	(%)
Stage		
pTa	114	25.1
pT1	340	74.9
Grade		
G1	260	57.3
G2	162	35.7
G3	32	7.0
Sex		
Men	383	84.4
Women	71	15.6
Number		
One	380	83.7
Two or more	74	16.3
Size		
$\leq$ 3 cm	357	78.6
$>$ 3 cm	97	21.4
Age		
$\leq$ 40 years	20	4.4
between 41 y 60 years	150	33
$>$ 61 years	284	62.6
Treatment		
Thiotepa	257	56.6
Adriamicine	33	7.3
Cisplatine	21	4.6
BCG	62	13.7
Others treatments	81	17.8

Table 1. Patients characteristics.

This methodology defines the following baseline survivor function:

$$S_0(t) = exp\,(-\sum_{j=1}^{k} \theta_j\, d_{ij}) \tag{3}$$

where $d_{ij} = 1$ if $t_j \leq t_i$, $d_{ij} = 0$ if $t_j > t_i$ and θ_j are given by:

$$\theta_j = \log \frac{S_0(t_{(j-1)})}{S_0(t_{(j)})} \tag{4}$$

Then it follows that the response probability can be expressed in the form:

$$p_i = 1 - exp\,(-exp\,(\beta' x_i) \sum_{j=1}^{k} \theta_j\, d_{ij}) \tag{5}$$

This leads to a *generalized non-linear model* for a binary response variables, with values y_i, and corresponding probabilities p_i, for $i = 1, 2, \ldots, n+c$. The model contains $k+p$ unknown parameters. After fitting the model, the statistic $-2 \log \hat{L}$ can be used to compare alternative manner.

Patients were followed up at clinic visits, generating observations as follows: 69 *left–censored*; 216 *right–censored* and the remaining patients are confined.

For the survivor function model, a minimal set of censoring times was chosen. The set of ordered censoring times is 50, 171, 261, 343, 399, 579, 674, 851, 1046, 1290, 1427, 1524, 1750, 2069, 2290, 2633, 2953, 3365, 3768, 5287.

We use the statistic $-2 \log \hat{L}$ in a strategy of selection of variables. We obtain number, tumor size and treatment as prognostic factors.

On fitting the model with tumor size and number of tumors the value of the statistic $-2 \log \hat{L}$ is 1498.4. On adding *Treatment* to the model, the value of this statistics is reduces to 1471.7. This reduction is significant at the 1% level.

Parameter	$\hat{\beta}$	Exp($\hat{\beta}$)	se($\hat{\beta}$)
two or more	0.2933	1.3408	0.1719
> 3 cm	0.3651	1.4406	0.1524
Cisplatine	0.3212	1.3787	0.2327
BCG	0.1279	1.1364	0.3314
ADR	0.6428	1.9017	0.1872
Others treatments	0.0733	1.0760	0.2027

Table 2. Generalized non-linear model. Parameters estimates

Using this model we may conclude that the relative hazard of first recurrence after TUR is increased in a 90% if adriamicine is provided, relative to a patient on thiotepa alone. The relative hazards are 1.37 and 1.13 respect thiotepa, when ciplastine and BCG are applied. This hazard is increased in a 7.3% if others treatments are applied, relative to a patient on thiotepa treatment alone. Patients with two or more tumors have a risk of recurrence 34% higher than patients with only one tumor and individuals with tumors > 3 cm have a risk 44% bigger than patients with tumor ≤ 3 cm.

We have checked the model by means of residuals proposed by Farrington in [Farrington, 2000]. It is assumed that the observation process that generates the interval censoring is independent of the survival times and the covariates. In that sense Figure 1 shows the distribution of interval lengths

by observation number. The plots do not reveal any systematic differences in the observation process between treatment groups.

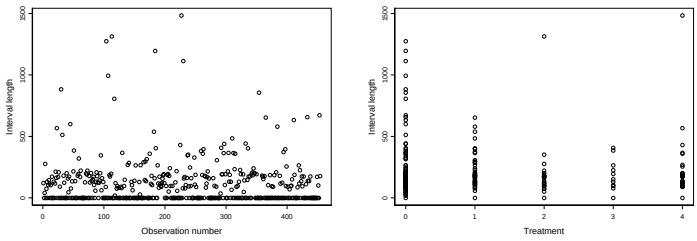

Fig. 1. Distribution of interval length: by observation number and treatment

Martingale residuals, in large samples, were shown to have zero mean under the correct model. That type of residuals reveal the existence of outliers. In Figure 2 patients 384 and 396 are separated from the bulk of the data. These patients belong to groups with the same features in size (> 3 cm), number (two or more) and treatment (Adriamicine).

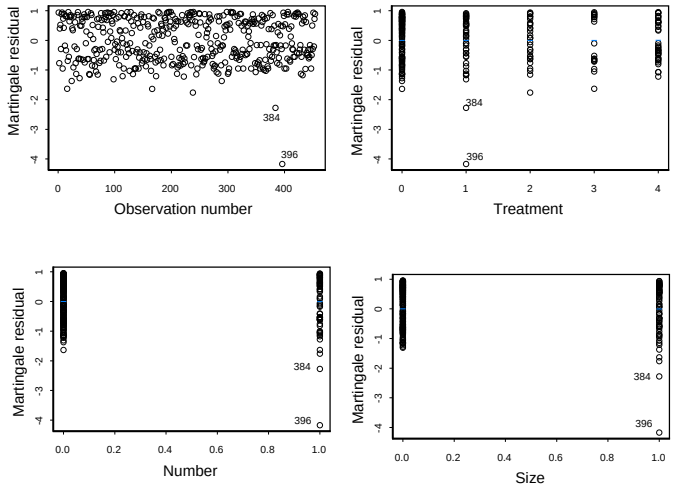

Fig. 2. Martingale residual by observation number, treatment, number and size

It would be useful to plot these residuals against log interval length and its analysis with deviance residuals as it is shown in [Farrington, 2000].

4 A Cox model of tumor recurrence

Let us consider now that time of recurrence is the time at which recurrence is detected.

The survival experience of the 454 patients depends on several variables, whose values have been recorded for each patient at the time origin. The aim of this Section is to determine which of explanatory variables have an impact on the free of disease time of the patients (survival time).

The focus is modelling the *recurrence hazard* (*risk of recurrence*)at time t. The *recurrence hazard* is obtained from the *hazard function* $h(t)$ and it is obtained from the basic model for survival data: *proportional hazard model* or *Cox regression model* given by:

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}) h_0(t)\,, \tag{6}$$

where $h_0(t)$ is the *baseline hazard function*.

On the other hand, the objective of this modelling procedure is to determine which combination of explanatory variables affects the form of the hazard function. In this process we use the statistic $-2Log\hat{L}$.

Indicator or *Dummies* variables are generated for the analysis. From *treatment* (five categories) four *Dummies* are defined: Adriamicine, Cisplatine, BCG and others treatments. From *grade* (three categories) two *Dummies*: G2 and G3. *Sex*, *number*, *size* and *stage* are dichotomic variables. *Age* is continuous. In this way the individual of reference is a 65 years old man (average patient), with only one tumor, of pTa stage, G1 grade, with a size minor or equal than 3 cm and with Thiotepa treatment after TUR.

Parameters estimates in the Cox regression model are presented in Table 3. The model allows us to compare risks among different groups of patients in a similar way of previous section.

Parameter	$\hat{\beta}$	Exp($\hat{\beta}$)	se($\hat{\beta}$)	**z**	**p-value**	**lower.95**	**upper.95**
> 3 cm	0.408	1.50	0.147	2.766	0.006	1.126	2.01
Cisplatine	0.418	1.52	0.224	1.866	0.062	0.979	2.36
BCG	0.201	1.22	0.329	0.611	0.540	0.642	2.33
ADR	0.725	2.07	0.181	4.017	0.000	1.450	2.94
Others treatments	0.147	1.16	0.201	0.731	0.460	0.781	1.72

Table 3. Cox regression model. Parameters estimates

Let us begin the model checking by testing the proportional hazards assumption. Grambsch and Therneau [Therneau and Grambsch, 2000] show

that the expected value of the ith *scaled Schoenfeld residual* is given by $E\left(r^*_{Pji}\right) \approx \hat{\beta}_j\left(t_i\right) - \hat{\beta}_j$, and so a plot of the values of $r^*_{Pji} + \hat{\beta}_j$ against the death times should give information about the form of the time-dependent coefficient of X_j, $\beta_j\left(t\right)$.

The horizontal line in each graph of Figure 3 indicates no suggestion of non-proportional hazards and that the coefficients of these variables are constant.

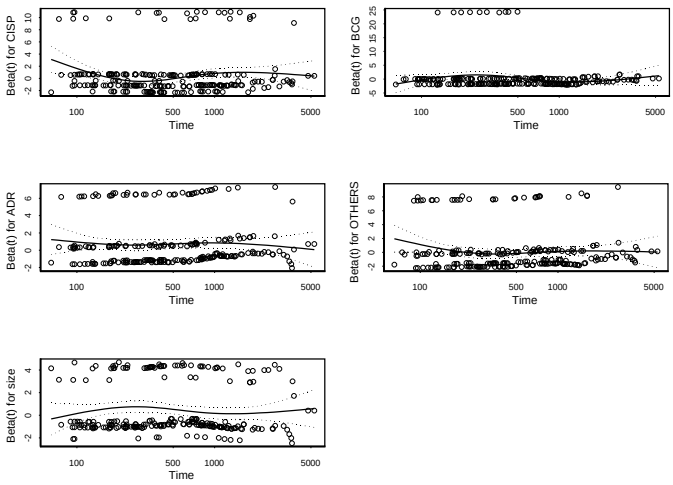

Fig. 3. Plots of scaled Schoenfeld residuals against time for each variable.

This graphical diagnostic is supplemented by a test for each variable, along with a global test for the model as a whole. In Table 4 it is showed the mentioned global test and the tests for each variable.

Here *rho* is the Pearson product-moment correlation between the *scaled Schoenfeld residuals* and time for each variable. The column *chisq* gives the tests statistics for each variable and the last row GLOBAL gives the global test for a χ^2 of 5 degree of freedom. With these results we may assume the proportional hazard hypothesis.

Validation and diagnostic of our model is based on *Martingale* and *Deviance* residuals. All results were consistent. The following graphics, see figure 4, show an Index plot of those residuals. In both plots the cluster of points is rather compact. We highlight patients 443 and 448 (they are patients 384 and 396 of section 3) whose survival times are larger than expected from the model.

variable	**rho**	**chisq**
Cisplatine	0.0182	0.0787
BCG	-0.0913	1.9992
Adriamicine	-0.0178	0.0748
Others treatments	-0.0524	0.6489
$\leq$ 3**cm**	-0.0352	0.2944
GLOBAL		**2.8889**

Table 4. Test for the Proportional Hazards

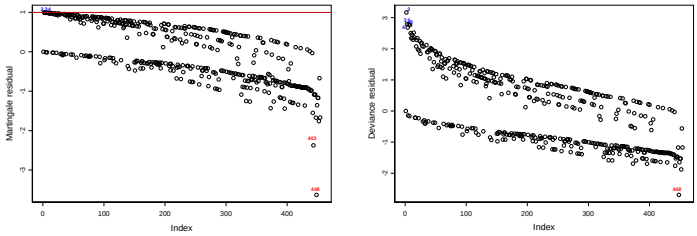

Fig. 4. Martingale and Deviance residuals

Identification of influential observations is performed by means of *Delta–Beta* test and examining the $-2\log\hat{L}$ changes. We found no alarming observations.

5 Conclusions

We have studied the prognostic factor for bladder cancer by means two different models: Cox regression and generalized non–linear models. In the first model, the prognostic factors are size and treatment; in the second model these factors are number, size and treatment. In the validation of both models the same two patients are detected and they belong to groups with the same features. Their characteristics correspond to the highest risk of recurrence and, however, they are among the patients with the longest time free of disease (what justify their behavior in our analysis). But this is not an important fact.

Acknowledgements: This work has been supported by the Generalitat Valenciana Grant GV04B-483

References

[Black *et al.*, 2002]R.J Black, R. Sankila, J. Ferlay, and D.M Parkin. Estimates of cancer incidence in europe for 1995. *Europe Journal Cancer*, pages 99–166, 2002.

[Collett, 2003]D. Collett. *Modelling Survival Data in Medical Research* 2^{th} *ed.* Chapman & Hall/CR, Boca Raton, Florida, 2003.

[Delwiche and Slaughter, 1998]L.D Delwiche and S.J Slaughter. *The Little SAS Book*. SAS Institute, Cary, NC, 1998.

[Farrington, 1996]C.P Farrington. Interval censored survival data: A generalized linear modelling approach. *Statistics in Medicine*, pages 283–292, 1996.

[Farrington, 2000]C.P. Farrington. Residuals for proportional hazards models with interval–censored survival data. *Biometrics*, pages 473–482, 2000.

[Hermanek and Sobin, 1998]P. Hermanek and L.H Sobin. *TNM Classification of malignant tumours* 4^{th} *ed.* Springer–Verlag, Berlin, 1998.

[Jaemal *et al.*, 2003]A. Jaemal, T. Murray, A. Samuels, E. Ghafoor, A. adn Ward, and M. Thun. Cancer statistics 2002. *CA Cancer J Clin*, pages 5–26, 2003.

[Kurth *et al.*, 1995]K.H Kurth, L. Denis, C. Bouffoux, R. Sylvester, F.M Debruyne, M. Pavone-Macaluso, and W. Oosterlinck. Factors affecting recurrence and progression in superficial bladder tumors. *Eug J Cancer*, pages 1840–46, 1995.

[Millan *et al.*, 2000]F. Millan, G. Chechile, J. Salvador, J. Palou, and J. Vicente. Multivariate analysis of the prognosis factors of primary superficial bladder cancer. *Journal Urology*, pages 73–78, 2000.

[Royston *et al.*, 2002]P. Royston, M. Parmar, and R. Sylvester. Flexible proportional–hazards and proportional–odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, pages 2175–2197, 2002.

[Therneau and Grambsch, 2000]T.M Therneau and P.M Grambsch. *Modeling Survival Data, Extending the Cox Model*. Springer, New York, 2000.

[Venables and Ripley, 2002]W.N Venables and B.D Ripley. *Modern Applied Statistics with S.* 4^{th} *ed.* Springer, New York, 2002.

[Zieger *et al.*, 1998]K. Zieger, H. Wolf, P.R Olsen, and K. Hojgaard. Long–term survival of patients with bladder tumours: the significance of risk factors. *Br. Journal Urology*, pages 667–72, 1998.