STATISTICAL TESTS TO COMPARE k SURVIVAL ANALYSIS FUNCTIONS INVOLVING RECURRENT EVENTS

Carlos Martínez [1,a], ***Guillermo Ramírez*** [2,b] & ***Maura Vásquez*** [2,c]
[1] *Associate Professor of Carabobo University (UC), Valencia, Venezuela,*
[2] *Titular Professor of the Central University of Venezuela (UCV), Caracas-Venezuela*

ABSTRACT

The objective of this paper is to propose statistical tests to compare k survival curves involving recurrent events. Recurrent events occur in many important scientific areas: psychology, bioengineering, medicine, physics, astronomy, biology, economics and so on. Such events are very common in the real world: viral diseases, seizure, carcinogenic tumors, fevers, machinery and equipment failures, births, murders, rain, industrial accidents, car accidents and so on. The idea is to generalize the weighted statistics used to compare survival curves in classical models. The estimation of the survival functions is based on a non-parametric model proposed by ***Peña*** *et al., using counting processes. R-language programs using known routines like survival and survrec were designed to make the calculations. The database* ***Byar*** *experiment is used and the time (months) of recurrence of tumors in 116 sick patients with superficial bladder cancer is measured. These patients were randomly allocated to the following treatments: placebo (47 patients), pyridoxine (31 patients) and thiotepa (38 patients). The aim is to compare the survival curves of the three groups and to determine if there are significant differences between treatments.*

Keywords: *Survival analysis, recurrent events, statistical tests.*

INTRODUCTION

Survival Analysis (***SA***) is a branch of Statistics which usually deals with death in biological organisms and failure in mechanical systems. It encompasses a wide variety of methods for analyzing the time to the occurrence of a given event. This event can be *death, disease, failure* or other event of interest. In most situations the occurrence of such event is undesirable, therefore, one of the primary goals is to evaluate the effect of some treatment on the prevention of such events. The classic theory of ***SA*** assumes that the event happens just once for each subject. Usually the analysis consist of estimating survival functions and comparing the survival distributions of several groups. However, many practical situations involve repeated events, where a subject or sample unit may experience any number of events over a lifetime. There is a growing interest in the analysis of this kind of events, known as recurrent events. When data include recurrent events, the statistical methods become more complicated. In this paper some recent developments and models will be discussed. (*See* **Andersen-Gill**[1] model and the proposes of **Peña** *et al*[21]).

BACKGROUND

The main random variable in ***SA*** is time of occurrence ***T.*** This variable has an associated set of useful functions: The probability density function (*pdf*) ***f,*** the cumulative distribution function (*cdf*) ***F,*** the survival function (*sf*) ***S***, the hazard function (*hf)* ***h*** and the cumulative risk function (*crf)* ***H***. By definition: ***F(t)=P(T≤t), S(t)=P(T>t)*** and consequently, ***S(t) = 1-F(t)***. Besides:

$$h(t)=\frac{f(t)}{S(t)} \quad and \quad H(t)=\int_0^t h(s)ds$$

Classical models in *SA*

Some of the classical models in ***SA*** are actuarial models (**Bhomer**[3], **Berkson-Gage**[2] and **Cutler-Ederer**[8]), **Kaplan-Meier**[14] (***KM***) and **Cox**[7] models. The estimate of the ***KM*** survival function is obtained as the product of conditional probabilities of occurrence of the event, where the observations are assumed independent. So the ***KM*** estimator with no survival times repeated is:

$$S(t_j)=\prod_{i=1}^{j}\left[1-\frac{d_i}{n_i}\right]$$

where ***j*** represents the ***j***th moment of occurrence, n_i the number of units at risk just before time t_j and d_i is the number of occurrence just in time t_j. **Cox**[7] proposed the proportional hazards model:

$$h(t/X)=h_0(t)\,e^{\beta' X} \quad and \quad S(t/X)=e^{-\int_0^t h(s/X)ds}$$

(a) E-mail: cmmm7031@gmail.com
(b) E-mail: guillermo.ramirez.ucv@gmail.com
(c) E-mail: mauralvasquez@gmail.com

where X is a vector of p covariates $X_1, X_2 \ldots X_p$, which are supposed to influence the time of occurrence. $h_o(t)$ is the baseline hazard function, $S(t/X)$ is the survival function and β the vector of parameters.

SA* Models with *RE

Wang-Chang[28] proposed a model for estimating of the common marginal survivor function in the case where interocurrence times are correlated. The estimator of **Wang-Chang** (***WC***) was defined using two processes: $d^*(t)$ which denotes the sum of the proportions of interoccurrence time equals t in the individuals, and $R^*(t)$ which represents the average number of individuals at risk at time t. Thus:

$$S(t) = \prod_{i=1}^{n} \prod_{\{j:T_{ij} \le t\}} \left[1 - \frac{d^*(T_{ij})}{R^*(T_{ij})} \right]$$

where T_{ij} denote the interocurrence time of the jth event in the ith subject (*see* ***Figure 1***).

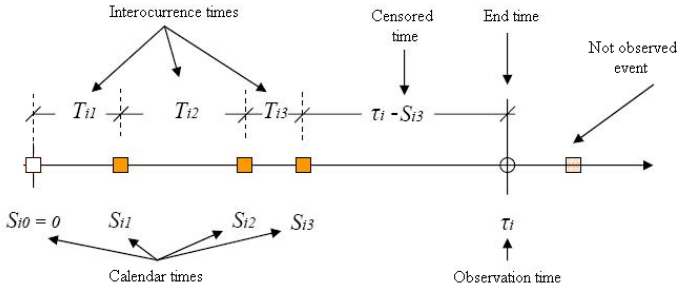

Figure 1. Pictorial representation of recurrent events in the* i*th *unit.

Peña *et al*.[21] proposed two survival models for **R**ecurrent **E**vents (***RE)***: One model called **GPLE** that generalizes the classical estimator of the ***KM*** survival function for units with ***RE***, assuming independent interoccurrence times; and the frailty model. In the first case it is necessary to consider two time scales: one related to calendar time S and the other to interoccurrence time T. The counting process $N(s,t)$ represents the number of observed events in the calendar period $[0,s]$ with $T \le t$ and $Y(s,t)$ represents the number of observed events in the period $[0,s]$ with $T \ge t$. The **GPLE** estimator is defined as:

$$S(t) = \prod_{\{w \le t\}} \left[1 - \frac{\Delta N(s,w)}{Y(s,w)} \right]$$

Where, $\Delta N(s,w) = N(s,w+\Delta w) - N(s,w)$ and Δw tends a cero.

Nonparametric methods for comparing groups in the SA

The aim of the comparison of groups in the ***SA*** is similar to procedures designed to compare independent samples test. However, there are certain aspects of survival analysis data, such as censoring and non-normality, that generate great difficulty when trying to analyze the data using these traditional statistical tests.

Several authors have developed statistical tests to compare groups in ***SA***, namely: the *logrank* test proposed by **Mantel-Haenszel**[18], generalized **Wilcoxon** test proposed by **Gehan**[11,12], **Mantel**[17] test, generalized **Kruskal-Wallis** test proposed by **Breslow**[5], **Cox**[7] test, **Peto-Peto**[23] test, **Tarone-Ware**[27] test, linear rank test with right censored data proposed by **Prentice**[24], **Harrington-Fleming**[13] test and a more general version given by **Fleming** *et al.*[10].

Next we describe the **Mantel-Haenszel**[18] test to compare two survival curves. Let us denote as d_{jz} the number of occurrences in time t_z in group j, for $j=1,2$ and $z = 1,2 \ldots p$. In the combined group $d_z = d_{1z} + d_{2z}$. ***Table 1*** illustrates that, in time t_z there are n_{1z} units at risk in the first group and n_{2z} units at risk in the second one. So, in that time t_z we have n_z units at risk in the combined group and d_z events will occur.

Table 1. Number of occurrences of the event at the time t_z for G1 and G2

Group	*Number of events*	*Number of survivors*	*Number at risk*
G1	d_{1z}	$n_{1z} - d_{1z}$	n_{1z}
G2	d_{2z}	$n_{2z} - d_{2z}$	n_{2z}
Pooled	d_z	$n_z - d_z$	n_z

The random variable d_{1z} follows a *hypergeometric distribution* $H(n_z,n_{1z},d_z)$ whose expected value is $d_z \times n_{1z}/n_z$ and whose variance:

$$Var(d_{1z}) = d_z \frac{n_{1z}}{n_z} \frac{n_z - n_{1z}}{n_z} \frac{n_z - d_z}{n_z - 1}$$

The null hypothesis of no difference between the survival curves of the groups can be tested using the statistics:

$$Z = \frac{\sum_{z=1}^{p} w_z \left[d_{1z} - E\left(d_{1z}\right) \right]}{\sqrt{\sum_{z=1}^{p} Var\left(d_{1z}\right)}}$$

where w_z are weight parameters for every moment t_z.

It can be shown that Z follows an asymptotic normal distribution and therefore Z^2 follows a χ^2 distribution with one degree of freedom. For different values of w_z several tests are obtained. If w_z=1 we have the **Mantel-Haenszel** test. For $w_z=n_z$ the generalized **Wilcoxon** test is attained. **Tarone-Ware**[27] proposed a modification of the **Wilcoxon** test with $w_z=(n_z)^{1/2}$. In **Peto-Peto**[23] test, $w_z = S_{PM}(t_z)$, where $S_{PM}(t_z)$ is an estimate of the survival function using the method of **Prentice-Marek**[25]. Other tests as **Fleming** *et al* [10] and **Harrington-Fleming**[13] have been adapted from previous tests. In the case of recurrent events, **Pepe-Cai**[22] indicated that the *logrank* comparison test can be adapted; **Doganaksoy-Nelson**[9] proposed a comparison method based on the mean cumulative function (**MCF**) to recurrent event data and **Martínez**[19] proposed a generalization of the classical weighted tests to recurrent events.

PROPOSAL

Problem

In this section we propose a statistic to test the hypothesis of no difference in k survival functions:

$$\mathbf{H_0}: S_1(t) = S_2(t) = \ldots = S_k(t)$$

where $S_1(t), S_2(t) \ldots S_k(t)$ are the survival functions of k groups of subjects under study.

Notation

The notation used is similar to the one used by **Peña** *et al.*[21], except for a third subscript r corresponding to the groups (*see* **Martinez** *et al.*[20]). A fundamental assumption of this approach is that individuals have been previously and properly classified in groups according to a stratification variable. Thus, we have to estimate a survival function for each group as:

$$\widehat{S}_r(t) = \prod_{z \le t} \left[1 - \frac{\Delta N(s,z;r)}{Y(s,z;r)} \right] \forall\ r = 1,2,\ldots,k.$$

Table 2 presents the summary of the data in the groups at the time of occurrence z, where $\Delta N(s,z)=\Delta N(s,z;1)+\ldots+\Delta N(s,z;r)+\Delta N(s,z;r')+\ldots+\Delta N(s,z;k)$ and $Y(s,z)=Y(s,z;1)+\ldots+Y(s,z;r)+Y(s,z;r')+\ldots+Y(s,z;k)$.The idea is to compare the proportion of occurrences in each group $\Delta N(s,z;r)/Y(s,z;r)$, to the respective proportion in the combined group $\Delta N(s,z)/Y(s,z)$.If H_o is true the proportion in the group is similar to the proportion in the combined group.

Table 2.Summary of the number of occurrences of the event in the groups in interocurrence time z

Group	*Number at risk*	*Number of events*	*Number of survivors*
1	*Y(s,z;1)*	*ΔN(s,z;1)*	*Y(s,z;1) - ΔN(s,z;1)*
.	.	.	.
r	*Y(s,z;r)*	*ΔN(s,z;r)*	*Y(s,z;r) - ΔN(s,z;r)*
r´	*Y(s,z;r')*	*ΔN(s,z;r')*	*Y(s,z;r') - ΔN(s,z;r')*
.	.	.	.
k	*Y(s,z;k)*	*ΔN(s,z;k)*	*Y(s,z;k) - ΔN(s,z;k)*
Pooled	*Y(s,z)*	*ΔN(s,z)*	*Y(s,z) - ΔN(s,z)*

The following statistic is a linear combination of the differences:

$$U_r = \sum_{z \le t} \left\{ Y(s,z;r) w_{rz} \left[\frac{\Delta N(s,z;r)}{Y(s,z;r)} - \frac{\Delta N(s,z)}{Y(s,z)} \right] \right\} \forall\ r = 1,2,\ldots,k.$$

where w_{rz} are the weights for each moment of occurrence z in the rth group. As the U_r variables satisfies $U_1 + U_2 + ... + U_k = 0$ with $w_{1z} = w_{2z} = ... = w_{kz} = w_z$, we can conclude that one of them is a linear combination of the others, and consequently, covariance matrix of U_r is not invertible.

$$Cov\{U_r, U_{r'}\} = \sum_{\forall z} \left\{ w_{rz}^2 \frac{Y(s,z) - \Delta N(s,z)}{[Y(s,z)-1]} \Delta N(s,z) \frac{Y(s,z;r)}{Y(s,z)} \left[\delta_{rr'} - \frac{Y(s,z;r)}{Y(s,z)} \right] \right\} \quad con\ \delta_{rr'} = \begin{cases} 1\ if\ r = r' \\ 0\ if\ r \neq r' \end{cases}$$

If one U_r variable is deleted, there is no substantial loss of information, and the covariance matrix of the new vector U with $k-1$ components is invertible, a required condition to propose the following statistic:

$$\chi^2_{gl=k-1} = U^t \Sigma_U^{-1} U$$

This statistic is a quadratic form and it will have an asymptotic chi-squared distribution with $k-1$ degrees of freedom under the null hypothesis. This expression is similar in structure to the one proposed by **Tarone-Ware**[27] to compare two groups. Choosing the appropriate weights, showed in ***Table 3***, we can obtain the corresponding generalization of some classical tests to the ***RE*** case. Note the subscript *rec* to indicate its adaptation to ***RE***.

Table 3. Proposed weights contrast tests for survival models with recurrence.

Test type	Test	Weight (w_z)	
Mantel-Haenzsel	**LR***rec*	*1*	Constant
Gehan	**G***rec*	*Y(s,z)*	Decreasing
Peto-Peto	**PP***rec*	$S_C(t)$	Decreasing
Tarone-Ware	**TW***rec*	$\sqrt{Y(s,z)}$	Decreasing
Peto-Prentice	**PPr***rec*	$S_C(t_{z-1})$	Decreasing
Prentice-Marek	**PM***rec*	$S_C(t) \times Y(s,z) / [Y(s,z) + 1]$	Decreasing
Fleming-Harrington-O'sullivan	**FHO***rec*	$[S_C(t)]^\rho$	$\rho \geq 1$: Decreasing $\rho = 0$: Constant $0 < \rho < 1$: Creasing
Fleming *et al.*	**F***rec*	$[S_C(t)]^\rho [1 - S_C(t)]^r$	-
Martinez	**M***rec*	$[S_C(t)]^\rho [1 - S_C(t)]^r Y(s,z)^\alpha / [Y(s,z) + 1]^\beta$	-
Martínez-Ramírez-Vásquez	**MRV***rec*	$\Delta N(s, z)$	Random

$Sc(t)$=*Estimate of survival function (**GPLE**) for the pooled group*

APPLICATION

The experiment data **Byar**[4] correspond to the time of recurrence (in months) of tumors in 116 patients with superficial bladder cancer (*see* **Martinez** *et al.*[20]). The patients were subjected to a process of randomization in the allocation of treatments: *placebo* (47 patients), *pyridoxine* (31 patients) and *thiotepa* (38 patients). At the beginning of the experiment, tumors in these patients were removed. In some cases, multiple tumors recurred were also removed to be found in the check-ups. Two variables were observed in each patient: the initial number of tumors (*num*) and size (*size*) or diameter of the largest, measured in centimeters.

The purpose of **Byar** investigation was to determine whether the variables num and/or size had significant effects on the recurrence of tumors, and then compare the three treatments. The objective in this application is to compare the survival curves of time to recurrence of tumors among the three groups: *placebo*, *thiotepa* and *pyridoxine*. Because one of the restrictions in the **GPLE** (**G**eneralized **P**roduct-**L**imit **E**stimator) model is that the latest time interocurrence is right censored, we had to make some changes to the database of patients.

DISCUSSION

Figure 2 illustrates the survival curves of the time of recurrence of tumors in the patients of the three groups. The proposed statistics were calculated and compared with results of previous investigations. In this example we used the estimates of survival functions with ***RE*** for different groups based on the statistic proposed by **Peña** *et al.*[21] (**GPLE**). In order to obtain the estimations some routines in **R language**[26] were written.

Table 4 illustrates that all tests indicated no significant differences among treatments except the **M***rec* test, with $\rho = 1$, $r = 1$, $\alpha = 1$ and $\beta = -1$.Observe that the tests can be obtained as particular cases of the **M***rec* except the **MRV**rec test. The tests most powerful are **M***rec*, **F***rec*, **LR***rec* and **MRV***rec*.

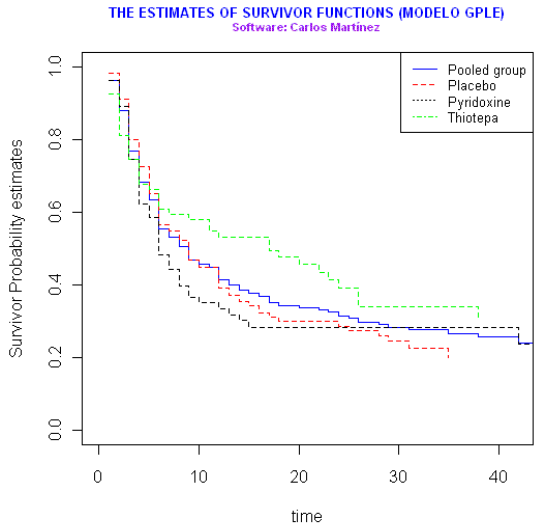

Figure 2. Pictorial representation of the comparisons of the three groups.

Table 4. Comparisons tests of the survival curves of the three groups.

Tests	**Chi-cuadrado**	**p-value**
LR*rec*	2.98493	0.22482
G*rec*	1.86934	0.39271
TW*rec*	2.37678	0.30471
PP*rec*	1.79447	0.40769
PM*rec*	1.78768	0.40908
PPr*rec*	1.95451	0.37634
F*rec* $\rho=0,\ r=0$	2.98493	0.22482
M*rec* $\rho=0, r=0,\ \alpha=0,\ \beta=0$	2.98493	0.22482
M*rec* $\rho=0, r=0,\ \alpha=1,\ \beta=0$	1.86935	0.39271
M*rec* $\rho=0, r=0,\ \alpha=1/2, \beta=0$	2.37678	0.30471
M*rec* $\rho=1, r=0,\ \alpha=0,\ \beta=0$	1.79447	0.40769
M*rec* $\rho=1, r=1,\ \alpha=1,\ \beta=-1$	6.67953	0.03544
MRV*rec*	3.16468	0.20549

REFERENCES

[1] **Andersen P., Gill, R.** (**1982**), Cox's regression model for counting processes: *A large sample study, Annals of Statistics*, 10, 1100-1120.

[2] **Berkson J.**, **Gage R.**, (**1952**), "Survival curve cancer patients following treatment". *American Statistical Association Journal*, 501-515.

[3] **Bhomer P.**, (**1912**), "Theorie der unabhängigen Wahrscheinlichkeiten" en Rapports, *Mémoires et Procès-verbaux de Septième Congrès International d´Áctuaries, Ámsterdam* vol. 2 p.p. 327-343.

[4] **Byar D.**, (**1980**): "The veterans administrations study of chemoprophylaxis for recurrent stage I bladder tumors: Comparisons of placebo, pyridoxine, and topical thiotepa. *In Bladder tumor and other Topics in Urological Oncology*. New York: Plenum, 363-370.

[5] **Breslow N.**, (**1970**): "A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika*, 57, 579-594.

[6] **Cochran W.**, (**1954**), "Some methods of strengthening the common $\chi 2$ tests", *Biometrics*, 10, 417-451.

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

[8] **Cutler S.**, **Ederer F.** (**1958**), Maximum utilization on the life table method in analyzing survival, Biostatistics, *Journal of chronic Diseases*, 8, 699-712.

[9] **Doganaksoy N.**, **Nelson W.**, (**1998**), "A method to Compare Two Samples of Recurrence Data", *Lifetime Data Analysis*, 48, 51-63.

[10] **Fleming T.**, **O'Fallon J., O'Brien P., Harrington D.** (**1980**), Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right-censored data, *Biometrics*, 36, 607-625.

[11] **Gehan E.**, (**1965a**), "A generalized Wilcoxon test for comparing arbitrarily singly-censored samples". *Biometrika*, 52, 203-223.

[12] **Gehan E.**, (**1965b**), "A generalized two-sample Wilcoxon test for doubly censored data". *Biometrika*, 52, 650-653.

[13] **Harrington D.**, **Fleming T.**, (**1982**), "A class of rank test procedure for censored survival data", *Biometrika*, 60 (3), 553-566.

[14] **Kaplan E.**, **Meier P.** (**1958**), Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association,* 53, 457-481.

[15] **Kruskal W.**, **Wallis W.,** (**1952**), "Use of ranks in one-criterion variance analysis". *Journal of the American Statistical Association,* 47, 583-621.

[16] **Mann H.**, **Whitney D.**, (**1947**), "On a test of whether one of two random variables is stochastically larger than the other". Ann. Math. Statist. 18, 50-60.

[17] **Mantel N.,** (**1966**), "Evaluation of survival data and two new rank order statistical arising in its consideration". *Cancer Chemother. Rep.*,50, 163-170.

[18] **Mantel N.**, **Haenszel W.**, (**1959**), "Statistical aspects of the analysis of data from retrospective studies of disease". *Journal of the National Cancer Institute*, 22 (4), 719-748.

[19] **Martínez C**., (**2009**), Generalizaciones de algunas pruebas clásicas de comparación de curvas de supervivencia al caso de eventos de naturaleza recurrente. *Tesis doctoral*. UCV, ISBN: 978-980-12-3605-4. Caracas-Venezuela.
[20] **Martínez C**., **Ramírez G**., **Vásquez M**. (**2009**), "Pruebas no paramétricas para comparar curvas de supervivencia de dos grupos que experimentan eventos recurrentes. Propuestas". *Revista ingeniería U.C*., Vol. 16, N° 3, 45-55.
[21] **Peña E., Strawderman R**., **Hollander M.** (**2001**), Nonparametric estimation with recurrent event data, *J.A.S.A*., 99, 1299-1315.
[22] **Pepe M**., **Cai J**., (**1993**), "Some graphical display and marginal regression analyses for recurrent failure times and time dependent covariates". *J.A.S.A*, 88, 811-820.
[23] **Peto R**., **Peto J**.,(**1972**), "Asymptotically efficient rank invariant test procedures (with discussion)". *J.R.S.S*., A, 135, 185-207.
[24] **Prentice R**. (**1978**), Linear rank tests with right censored data, *Biometrika*, 65(1), 167-179.
[25] **Prentice R**., **Marek, P**. (**1979**), Qualitative discrepancy between censored data rank test, *Biometrics*, 35, 861-867.
[26] **R Development Core Team** (**2010**). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
[27] **Tarone R**., **Ware J**., (**1977**), "On distribution-free tests for equality of survival distributions". *Biometrika*, 64 (1), 156-160.
[28] **Wang M**., **Chang S**. (**1999**), Nonparametric estimation of a recurrent survival function, *Journal of the American Statistical Association*, 94, 146-153.
[29] **Wilcoxon F**., (**1945**), "Individual comparisons by ranking methods". *Biometrics*, 1, 80-83.

ACKNOWLEDGEMENTS

We are most grateful to Aleida Aular, Rafael Graterol and Silvio Pérez for their invaluable collaboration in providing us with comments and suggestions which led to approval of this paper.