Competing risks survival analysis with recurrent events

Carlos M. Martínez M.
University of Carabobo, Venezuela, cmmm7031@gmail.com

Abstracts

The competing risks too arise when one type of event may affect the probability of occurrence of other events. Some authors have made important contributions in this area of research, multistate multivariate models have been proposed. Varieties of record are frequently encountered in areas as *medicine, engineering, sociology, biology, social science*, among others. Classical survival analysis models in competing risks of multiple non-recurrent single events are possible to generalize to multiple recurrent events. This paper presents and reviews the formulation of a competing risks model with recurrent events and standard methods. The objective of this paper is to propose a model in competing risks to recurrent events. The data in the models in classical competing risks usually arises in studies in which the *k* events of a subject are considered as independent, mutually exclusive and not recurrent events, with ($k>1$). This work also illustrated the methodology to apply the proposed model and an example with simulated data is developed. The estimations of the methods presented here are calculated using the **R** package. Counting processes are used. Some routines of **R** language are show and used to estimate functions of the model in competing risks with recurrent events.

Key Words: Competing risks, counting process, survival analysis, recurrent events

1. Introduction

Survival analysis is a collection of statistical tools and methods for inference on time to event data. When the data consist of patients who experience only one type of event and censored individuals, a classical nonparametric estimate of the cumulative incidence can be obtained using **actuarial methods**, **Kaplan-Meier estimator** or **Nelson-Aalen method**. If, we need assess the effect of covariates on the occurrence of a single event, another traditional method can be used. The cumulative incidence can be obtained using traditional **Cox model**. If the study is oriented to study the occurrence one single recurrent event, we can use different estimators as **Wei** *et al.* **1981**, **Andersen** and **Gill 1982**, **Prentice** *et al.* **1989**, **Wang** and **Chang 1999**, **Peña** *et al.* **2001**, frailty model or dynamic model of **Peña**. R packages for the analysis of time to event data have been designed. The *survival* package calculates both the Kaplan-Meier estimator and Cox type estimator in traditional survival analysis. The *muhaz* package permits to estimate the hazard function through kernel methods for right-censored data. The *survrec* package too proposes implementations of several models for recurrent events data, such as the **Peña** *et al*., **Wang** and **Chang** estimators, and **MLE** (**M**aximum **L**ikelihood **E**stimator) estimation under a frailty model. Competing risks arise in studies when subjects are exposed to more than one cause of failure and failure due to one cause excludes failure due to other causes. For competing risks data, a useful quantity is the cumulative incidence function, which is the cumulative probability of one event from a type of specific cause over time (*see* **Pintilie 2006**). Traditional nonparametric methods in competing risks analysis can be used for the analysis the causes on the occurrence of an event or classical regression methods for survival analysis too can be based directly for estimate the cause-specific hazard function or on the cumulative incidence function. **R** too provides programs for competing risks. *mstate* or *cmprsk* are package that are used to estimate hazards and probabilities, in the context of competing risks and multistate models.

2. Background

2.1 Classical survival analysis

Standard survival data measure the time span from some time origin until the occurrence of one type of event. Standard survival analysis considers the time until some first event only.

The **figure 01** shows a pictorial representation of a graphic model of classical study of survival analysis.

Figure 01 Representation graphics of the classical survival data situation.

To describe the estimator of survival analysis is need to introduce some notation. Let ***T*** be a continuous random variable representing survival time, ***δ*** denote a ***(1,0)*** random variable indicating either the occurred event or censorship. That is, ***δ=1*** for event if occurs during the study or ***δ=0*** if the survival time is censored. ***S(t)*** is the survivor function, ***h(t)*** is hazard function and ***H(t)*** accumulative hazard function.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t / T \geq t)}{\Delta t}$$

Its demonstrate that,

$$h(t) = \frac{f(t)}{S(t)} \qquad S(t) = e^{-\int_0^t h(s)ds} \qquad H(t) = \int_0^t h(s)d(s)$$

2.2 Survival Analysis for one single recurrent event

When each person can experience more than one event, the event of interest occurs repeatedly in the same subject. The analysis is known as recurrent events. Recurrent event data arise in diverse fields such as *medicine, public health, insurance, social science, economics, manufacturing* and *reliability*.

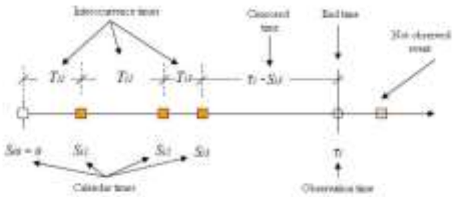


Figure 02 Representation graphics of the survival analysis with recurrent event.

The regression methods and models type Cox for recurrent events were studied by **Prentice** *et al*. **1981**, **Andersen** and **Gill 1982** and **Wei** *et al*. **1989** and nonparametric models were studied by **Wang** and **Chang 1999** and **Peña** *et al*. **2001.**

2.2.1 Wang-Chang model (WC model)

If data are correlated, the model that should be used is the estimator proposed by **Wang** and **Chang** (**1999**) that take into account the fact that interocurrence times were or not correlated. Wang and Chang proposed an estimator for marginal survival function in case there is correlation between the times of occurrences. They consider correlation structure generally includes as a special case of independence. Your estimator can be defined using two processes, ***d**** and ***R****. The authors attempt to take into account in the definition that an individual may have more than one event. The model,

$$\hat{S}(t) = \prod_{T_{k \leq t}} \left[1 - \frac{d^*(T_k)}{R^*(T_k)} \right]$$

2.2.2 PHS model

If data are independents, the interocurrence survival function can be estimated by **Peña** *et al.* (**2001**). They generalize the product limit estimator of Kaplan-Meier curves for the case with recurrent events processes using counters. This formulation defines two functions, ***N*** and ***Y***, which in the particular case with a single occurrence survival for individual, indicating the number of events occurring in a period of time and individuals at risk for a given time, respectively.

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

The counters processes ***N*** and ***Y*** are defined simultaneously in two time scales (calendar time scale and the time scale of intercurrence). Thus, ***N(s, t)*** is defined as the number of observed events occurring for a time period ***[0, s]*** whose time intercurrence times are at most ***t***, and ***Y(s,t)*** is defined as observed number of events on calendar time ***[0, s]*** whose intercurrence times are greater and equal to ***t***.

2.3 Traditional competing risks models

If several types of events (*says k events*) occur, a model describing progression to each of these competing risks is needed. The **figure *03*** shows a pictorial representation of a graphic model of competing risks. Recent techniques have been developed by **Gray 1988** and **Fine** and **Gray 1999** take into account the competing risks. Competing risks are said to be present when a patient is at risk of more than one mutually exclusive event.

Competing risks model is specified through the characterization of the joint distribution of (T,r,δ). The distinct specifications of the survival function will be focus in the competing risks framework. Define, for each ***ith*** unit, (T_i, r_i, δ_i), T is survival time as $T_i = \min\{Y_i, C_i, \tau_i\}$ where Y_i is the time of occur of the event in the ***ith*** unit, C_i is censoring time of i*th* unit and τ_i is observation time of unit, j is the type of event and δ_i is variable of censoring. T is assumed to be a continuous and positive random variable, and r_i takes values in the finite set $\{1,2,\ldots,k\}$ and it represent the type of event. $\delta_i = 0$ if the data is censored and if $\delta_i = 1$ occurred a type of event. It is considered that the units experiments from one and only one type event. The joint distribution of (T,r,δ) is completely specified through either the cause-specific hazards or through the cumulative incidence functions $F_j(t)$.

Figure 03 Pictorial representation of a graphic model of competing risks.

The cause-specific hazard function for the ***jth*** event is defined as

$$h_j(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t / T \ge t, r = j)}{\Delta t} \quad \forall\, j = 1,2,\ldots,p$$

$h_j(t)$ represents the rate of occurrence of the ***jth*** event. The cumulative incidence function from type ***jth*** event, also referred to as sub distribution function, is defined by

$$F_j(t) = P(T \le t, r = j) \ \ \forall\, j = 1,2,\ldots,p$$

and corresponds to the probability of an unit experiments the j*th* event in the presence of all the competing risks. The distribution function for **T** is obtained from the cumulative incidence functions through

$$F(t) = \sum_{j=1}^{p} F_{j(t)}$$

The total hazard h(t) is defined as follow

$$h(t) = \sum_{j=1}^{p} h_{j(t)}$$

and the overall survival function ***S(t)*** for ***T*** is defined in terms of the specific hazards as follows:

$$S(t) = e^{-\sum_{j=1}^{p}\int_0^t h_j(s)ds}$$

$$S(t) = \prod_{j=1}^{p} e^{-\int_0^t h_j(s)ds}$$

So, the overall survival function ***S(t)*** is obtained from the relationship

$$S(t) = \prod_{j=1}^{p} S_j^*(t)$$

Where,

$$S_j^*(t) = e^{-\int_0^t h_j(s)ds}$$

These are call incidence functions.

3.1 Nonparametric model type Kaplan-Meier to competing risks

The Kaplan-Meier estimator can easily be generalized to include competing risks. Let

$$t_{j1} \leq t_{j2} \leq \ldots \leq t_{jk_j}$$

denote the k_j distinct failure times for failures of type j. Let n_{ji} denote the number of subjects at risk just before t_{ji} and let d_{ji} denote the number of deaths due to cause j at time t_{ji}. Then the same arguments used to derive the traditional Kaplan-Meier estimator lead to,

$$\hat{S}_j(t) = \prod_{t_{ji} \leq t} \left[1 - \frac{d_{ji}}{n_{ji}}\right]$$

3.3 Cox proportional hazard regression model to competing risks

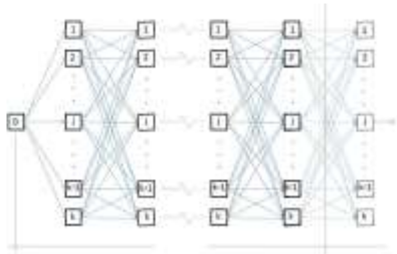
Cox model was proposed by **Lunn** and **McNeil** (**1995**). **Fine** and **Gray (1999)** developed a method for regression analysis that models the hazard that corresponds to the cumulative incidence function. Cox proportional hazard model can be used to estimate the covariate effects on hazard function for each type of event. If the proportional hazard assumption does not hold across the causes, the stratified analysis should be used, that is equivalent of fitting separate model for each event type. The specific hazard functions are given by¸

$$\lambda_j(t; x) = \lambda_{0j}(t)\, exp(\beta^t_j X)$$

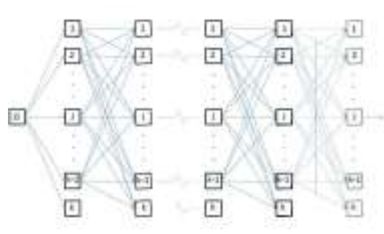
Where $\lambda_{0j}(x)$ is the baseline of the specific hazard for the event j*th* and β_j is the vector of the regression coefficients related to the event j*th*.

4. Proposal

The objective of this paper is to propose a model in competing risks to recurrent events. **Figure 04** shows a pictorial representation of two graphic models to competing risks with recurrent events within and with a terminal event.



(a)



(b)

Figure 04 Pictorial representations of two graphic models of competing risks with recurrent events.

The figure 4.a shows the recurrence of *k* events on time without terminal event. While that figure 4.b shows the recurrence of *k-1* events on time with presence of an event terminal. The complexity of the models depends on three fundamental assumptions. First, that the times between events intercurrence are uncorrelated. Second, the events are mutually exclusive. At a certain time, will not occur more than one event in the same unit study. And finally, that the terminal event may or may not occur in the unit. This paper covers in competing risks problem with recurrence of *k* mutually exclusive types of events, without the presence of a terminal event (*see figure 4.a*). The idea is to generalize the **Peña** *et al.* (**2001**) and **Wang** and **Chang** (**1999**) models to the case of *k* competing risks for recurrent event types. If intercurrence times are assumed mutually independent PHS model can be used. If times are correlated intercurrence **WC** model can be used. So,

For independent intercurrence times:

$$\hat{S}_j(s,t) = \prod_{w \leq t} \left[1 - \frac{\Delta N_j(s,w)}{Y(s,w)}\right]$$

For correlated intercurrence times:

$$\hat{S}_j(t) = \prod_{T_k \leq t} \left[1 - \frac{d_j^*(T_k)}{R^*(T_k)}\right]$$

Where, $S_j(s,t)$ and $S_j(t)$ are the survival functions for each event-specific j*th* in the competing risks. Both models can be used to estimate functions of competing risks in biomedical researches when subjects are at risk of occurrence from *k* different recurrent events with independent data or correlated data, respectively.

5. Applications

For illustration of the competing risks models with recurrent events, simulated data will be used. In this paper, it is will be used data of two hypothetical cases, with the presence of competing risks and with two recurrent events. In the first case, it is assumed that the study unit experienced only one type of event of the two types of recurrent events that the unit can experience. In this case, there are 539 units. To these units, it is have generated data of occurrence of two events in competing risks. **Table 01** contains a sample of the data. In the second case, it is assumed that the unit may experience both types of events. Too, it was used 539 units. **Table 02** shows part of simulated data for this case.

Table 01 Each unit experienced only one type of event.

Obs	Unit	Gap time	Type	Censored
530	505	19.90417522	1	0
531	506	20.32854209	1	0
532	507	20.33675565	1	0
533	508	20.67898700	1	0
534	509	20.72005476	2	0
535	510	20.75564682	2	0
536	**511**	**21.26488706**	**2**	**1**
537	**511**	**21.26488706**	**2**	**1**
538	**511**	**21.26488706**	**2**	**0**
539	512	21.84804928	1	0

Table 02 Each unit can experienced both type of events.

Obs	unit	Gap time	type	Censored
300	**281**	**5.908281999**	**1**	**1**
301	**281**	**5.908281999**	**2**	**1**
302	**281**	**5.908281999**	**2**	**1**
303	**281**	**5.908281999**	**2**	**0**
304	282	5.938398357	1	0
305	283	5.938398357	1	0
306	284	5.94934976	1	0
307	285	5.952087611	2	0
308	286	5.976728268	1	0
309	287	6.045174538	1	0

6. Results

Results of the applications methods designed for the competing risks analysis with recurrent events are shows on the **tables 03** and **04**, respectively. The method of estimation used is **PHS model**. We were assumed independence in the data. But, we can use **WC model** on case of correlated data. Estimates of the functions of survival are obtained using routines R package.

Table 03 Results of the application of the PHS model where each unit experienced only one type of event

Time	Type	Survival	$\Delta N(s,t)$	$Y(s,t)$	$\lambda(s,t)$	$1-\lambda(s,t)$	$S(s,t)$	$\Delta N_1(s,t)$	$\lambda_1(s,t)$	$1-\lambda_1(s,t)$	$S_1(s,t)$	$\Delta N_2(s,t)$	$\lambda_2(s,t)$	$1-\lambda_2(s,t)$	$S_2(s,t)$	$S^*(s,t)=S_1(s,t)*S_2(s,t)$
0.0000	-	1.0000	0	539	0.00000	1.00000	1.0000	0	0.00000	1.00000	1.0000	0	0.00000	1.00000	1.00000	1.0000
0.5585	1	0.9961	2	518	0.00386	0.99614	0.9961	2	0.00386	0.99614	0.9961	0	0.00000	1.00000	1.00000	0.9961
0.9884	1	0.9920	2	486	0.00412	0.99588	0.9920	2	0.00412	0.99588	0.9920	0	0.00000	1.00000	1.00000	0.9920
1.4182	1	0.9877	2	456	0.00439	0.99561	0.9877	2	0.00439	0.99561	0.9877	0	0.00000	1.00000	1.00000	0.9877
1.9630	1	0.9807	3	425	0.00706	0.99294	0.9807	3	0.00706	0.99294	0.9807	0	0.00000	1.00000	1.00000	0.9807
2.6968	2	0.9732	3	389	0.00771	0.99229	0.9732	0	0.00000	1.00000	0.9807	3	0.00771	0.99229	0.99229	0.9732
3.4524	1	0.9650	3	358	0.00838	0.99162	0.9650	3	0.00838	0.99162	0.9725	0	0.00000	1.00000	0.99229	0.9650
4.8569	1	0.9586	2	304	0.00658	0.99342	0.9586	2	0.00658	0.99342	0.9661	0	0.00000	1.00000	0.99229	0.9586
5.9083	2	0.9516	2	271	0.00738	0.99262	0.9516	0	0.00000	1.00000	0.9661	2	0.00738	0.99262	0.98496	0.9516
7.0828	2	0.9430	2	221	0.00905	0.99095	0.9430	0	0.00000	1.00000	0.9661	2	0.00905	0.99095	0.97605	0.9430
9.6071	2	0.9314	2	163	0.01227	0.98773	0.9314	0	0.00000	1.00000	0.9661	2	0.01227	0.98773	0.96407	0.9314
11.7317	2	0.9168	2	128	0.01563	0.98438	0.9168	0	0.00000	1.00000	0.9661	2	0.01563	0.98438	0.94901	0.9168
21.2649	2	0.8613	2	33	0.06061	0.93939	0.8613	0	0.00000	1.00000	0.9661	2	0.06061	0.93939	0.89150	0.8613
29.2567	2	0.6460	1	4	0.25000	0.75000	0.6460	0	0.00000	1.00000	0.9661	1	0.25000	0.75000	0.66862	0.6460
29.6674	1	0.3230	1	2	0.50000	0.50000	0.3230	1	0.50000	0.50000	0.4831	0	0.00000	1.00000	0.66862	0.3230

Table 04 Results of the application of the PHS model where each unit can experience both events.

Time (t)	AtRisk	$\Delta N(s,t)$	$Y(s,t)$	$\lambda(s,t)$	$1-\lambda(s,t)$	$S(s,t)$	$\Delta N_1(s,t)$	$\lambda_1(s,t)$	$1-\lambda_1(s,t)$	$S_1(s,t)$	$\Delta N_2(s,t)$	$Y(s,t)$	$\lambda_2(s,t)$	$1-\lambda_2(s,t)$	$S_2(s,t)$	$S^*(s,t)=S_1(s,t)*S_2(s,t)$
0.0000	539	0	539	0.00000	1.00000	**1.000000**	0	0.00000	1.00000	**1.00000**	0	539	0.00000	1.00000	**1.000000**	**1.0000000**
0.5585	521	3	521	0.00576	0.99424	**0.994242**	2	0.00384	0.99616	**0.99616**	1	521	0.00192	0.99808	**0.998081**	**0.9942492**
0.9884	488	3	488	0.00615	0.99385	**0.988130**	2	0.00410	0.99590	**0.99208**	1	488	0.00205	0.99795	**0.996035**	**0.9881454**
1.4182	457	2	457	0.00438	0.99562	**0.983805**	2	0.00438	0.99562	**0.98774**	0	457	0.00000	1.00000	**0.996035**	**0.9838209**
1.9630	426	3	426	0.00704	0.99296	**0.976877**	3	0.00704	0.99296	**0.98078**	0	426	0.00000	1.00000	**0.996035**	**0.9768926**
2.6968	390	3	390	0.00769	0.99231	**0.969363**	0	0.00000	1.00000	**0.98078**	3	390	0.00769	0.99231	**0.988374**	**0.9693780**
3.4524	359	3	359	0.00836	0.99164	**0.961262**	3	0.00836	0.99164	**0.97259**	0	359	0.00000	1.00000	**0.988374**	**0.9612774**
4.8569	305	2	305	0.00656	0.99344	**0.954959**	2	0.00656	0.99344	**0.96621**	0	305	0.00000	1.00000	**0.988374**	**0.9549739**
5.9083	272	3	272	0.01103	0.98897	**0.944426**	1	0.00368	0.99632	**0.96266**	2	272	0.00735	0.99265	**0.981106**	**0.9444669**
7.0828	221	2	221	0.00905	0.99095	**0.935879**	0	0.00000	1.00000	**0.96266**	2	221	0.00905	0.99095	**0.972227**	**0.9359197**
9.6071	163	2	163	0.01227	0.98773	**0.924396**	0	0.00000	1.00000	**0.96266**	2	163	0.01227	0.98773	**0.960298**	**0.9244360**
11.7317	128	2	128	0.01563	0.98438	**0.909952**	0	0.00000	1.00000	**0.96266**	2	128	0.01563	0.98438	**0.945293**	**0.9099917**
21.2649	33	2	33	0.06061	0.93939	**0.854804**	0	0.00000	1.00000	**0.96266**	2	33	0.06061	0.93939	**0.888003**	**0.8548407**
29.2567	4	1	4	0.25000	0.75000	**0.641103**	0	0.00000	1.00000	**0.96266**	1	4	0.25000	0.75000	**0.666002**	**0.6411305**
29.6674	2	1	2	0.50000	0.50000	**0.320551**	1	0.50000	0.50000	**0.48133**	0	2	0.00000	1.00000	**0.666002**	**0.3205653**

5. Discussions of results and conclusions

The competing risks models nonparametric for recurrent events are proposed in this paper. Of our example, we can deduce that,

$$\lambda(s,t)=\lambda_1(s,t)+\lambda_2(s,t)+\ldots+\lambda_k(s,t)$$

The above equality is satisfied if the types of events are mutually exclusive on time and where each unit experienced only one type of event. So,

$$\hat{S}(s,t)=\prod_{w\leq t}\left[1-\sum_{j=1}^{k}\frac{\Delta N_j(s,w)}{Y(s,w)}\right] \quad and \ \hat{S}(s,t)=\prod_{j=1}^{k}S_j^*(s,t) \ with \ S_j^*(s,t)=\prod_{w\leq t}\left[1-\frac{\Delta N_j(s,w)}{Y(s,w)}\right]$$

Now, if the events are mutually exclusive on time and each unit can experience the *k* type of event. For the law of probabilities, we can say that the total hazard ***λ(s,t)*** and the overall survival function ***S(s,t)*** are defined as,

$$\lambda(s,t)=\sum_{j=1}^{k}\lambda_j(s,t)-\sum_{j'<j=1}^{k}\lambda_j(s,t)\lambda_{j'}(s,t)+\sum_{j''<j'<j=1}^{k}\lambda_j(s,t)\lambda_{j'}(s,t)\lambda_{j''}(s,t)+\cdots+(-1)^{k+1}\prod_{j=1}^{k}\lambda_j(s,t)$$

And,

$$\hat{S}(s,t)=\prod_{w\leq t}\left[1-\sum_{j=1}^{k}\frac{\Delta N_j(s,w)}{Y(s,w)}+\sum_{j'<j=1}^{k}\frac{\Delta N_j(s,w)\Delta N_{j'}(s,w)}{[Y(s,t)]^2}-\sum_{j'<j=1}^{k}\frac{\Delta N_j(s,w)\Delta N_{j'}(s,w)\Delta N_{j''}(s,w)}{[Y(s,t)]^3}+\cdots+(-1)^{k+1}\frac{\prod_{j=1}^{k}\Delta N_j(s,w)}{[Y(s,t)]^k}\right]$$

So,

$$\hat{S}(s,t)=\prod_{w\leq t}[1-\lambda(s,t)] \quad and \ \hat{S}(s,t)=\prod_{j=1}^{k}S_j^*(s,t) \ with \ S_j^*(s,t)=\prod_{w\leq t}\left[1-\frac{\Delta N_j(s,w)}{Y(s,w)}\right]$$

On this work, we were present and show important models that are need for the analysis of competing risks data with recurrent events.

References

Andersen, P.K., Gill, R.D. (1982). Cox's regression model for counting processes – a large sample study. Ann. Statist. 10, 1100-1120.

Fine, J.P., Gray, R.J. (1999). A proportional hazards model for the subdistribution of a competing risk. J. Amer. Statist. Assoc. 94, 496-509.

Gray, R.J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. Ann. Statist. 16, 1141-1154.

Lunn M, McNeil D (1995). Applying Cox Regression to Competing Risks. Biometrics, 51, 524-532.

Peña, E., Strawderman, R. y Hollander, M. (2001). Nonparametric estimation with recurrent event data. J.A.S.A., 99, 1299-1315.

Pintilie, M. (2006), Competing Risks: A Practical Perspective, Wiley.

Prentice, R., Williams, B. y Peterson, A. (1981). On the regression analysis of multivariate failure time data. Biometrika, 68(2), 373-379.

Putter H, Fiocco M, Geskus R (2007). Tutorial in Biostatistics: Competing Risks and Multi-State Models. Statistics in Medicine, 26(11), 2389-2430.

Wang, M. y Chang, S. (1999). Nonparametric estimation of a recurrent survival function. Journal of the American Statistical Association, 94, 146-153.

Wei, L., Lin, D. y Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association, 84, 1065-1073.