

Instituto Nacional de Angiología y Cirugía Vascular

## **LA MODELACIÓN DE DATOS DE SUPERVIVENCIA EN ANGIOLOGÍA: EL ANÁLISIS DE LAS INTERACCIONES**

*Dr. Armando H. Seuc<sup>1</sup>, Dr. Rafael Simón<sup>2</sup>, Dr. Emma Domínguez Alonso<sup>3</sup>  
y Dr. Lázaro Chirino<sup>4</sup>*

**RESUMEN:** Se presenta el modelo de riesgos proporcionales (conocido también como modelo de regresión de Cox) para el análisis de datos de supervivencia. Se introduce el concepto de interacción entre 2 factores de este modelo, y se discuten las opciones para la presentación e interpretación de los resultados. Los procedimientos presentados se ilustran mediante un ejemplo en el que se modela el tiempo hasta la reoclusión en pacientes revascularizados del sector aortofemoral, a partir de algunas variables predictoras, en particular la lipoproteína(a) y el LDL-colesterol.

**DeCS: MODELOS DE RIESGOS PROPORCIONALES; ANALISIS DE SUPERVIVENCIA; BIOMETRIA**

El modelo básico para el análisis de datos de supervivencia es el modelo de riesgos proporcionales (conocido también como modelo de regresión de Cox) propuesto por Sir David Cox en 1972.<sup>1</sup> En este modelo el interés radica en la función de riesgo instantáneo de “morir”  $h(t)$ , denominada “*hazard function*” en la literatura de lengua inglesa, y que se define como la probabilidad de que un individuo “muera” (o le ocurra el evento de interés, ya sea positivo o negativo) al tiempo  $t$ , dado que ha sobrevivido hasta ese mismo momento.

Se puede demostrar que la función de riesgo  $h(t)$  es:

$$h(t) = d/dt [\log S(t)],$$

donde  $S(t)$ , la función de supervivencia, se define como

$$S(t) = P(T \geq t),$$

siendo  $T$  la variable aleatoria “tiempo de supervivencia”.

<sup>1</sup> Doctor en Ciencias Matemáticas. Investigador Auxiliar.

<sup>2</sup> Investigador Auxiliar. Doctor en Ciencias de la Salud.

<sup>3</sup> Especialista de I Grado en Bioestadística.

<sup>4</sup> Residente de Angiología

Las funciones de riesgo y de supervivencia  $h(t)$  y  $S(t)$  respectivamente, son las que acaparan generalmente la atención en el análisis de datos de supervivencia; esto representa una diferencia importante con respecto a otros modelos lineales (regresión y ANOVA) en los que se modela la dependencia de la respuesta promedio (o una función de ella) en una cierta variable “diana”, con respecto a un grupo de variables predictoras. La razón fundamental de esta singularidad es que la frecuente presencia de datos censurados y la usual asimetría de los datos de supervivencia hace inadecuado e impracticable el cálculo de la supervivencia promedio en un conjunto de datos de este tipo.

De todas maneras la modelación de la función de riesgo permite, indirectamente, la estimación de una medida de tendencia central como la mediana del tiempo de supervivencia, ya que a partir de la función de riesgo se puede estimar la función de supervivencia y, a partir de esta última, la mediana del tiempo de supervivencia.

En el modelo de riesgos proporcionales la función de riesgo para el  $i$ -ésimo individuo se define como:

$$h_i(t) = y(x_i) h_0(t),$$

donde:

$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  es el vector con los valores para el  $i$ -ésimo individuo en las  $p$  variables  $X_1, X_2$  y  $X_p$ , las que se suponen predictoras de la función de riesgo. Se supone que la medición de estas  $p$  variables ocurre al inicio de la observación (el tiempo cero) en cada individuo.

$h_0(t)$  es la función de riesgo basal; es la función de riesgo para un individuo en el cual todas las  $p$  variables predictoras toman el valor cero.

$\psi(\cdot)$  es una función que puede interpretarse como la razón entre el riesgo (instantáneo de morir) de un individuo con vector observado de variables predictoras  $x_i$  y el riesgo (instantáneo de morir) de un individuo con vector observado de variables predictoras  $x = 0$ . Como esta función no depende del tiempo, se está suponiendo que esta razón es la misma para todo  $t$ .

Como el riesgo (instantáneo de morir) relativo  $\psi(\cdot)$  es siempre no negativo, podemos considerar  $y(x_i) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})$ ,

por lo que el modelo de riesgos proporcionales puede escribirse como  $h_i(t) = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) h_0(t)$ , es decir,  $\log[h_i(t)/h_0(t)] = (\beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})$ ,

de ahí que el modelo de riesgos proporcionales puede considerarse como un modelo lineal para que el logaritmo de la razón de riesgos (instantáneos de morir), lo cual lo inserta en la categoría de los modelos lineales generalizados.<sup>2</sup>

### **La inclusión de interacciones en el modelo de riesgos proporcionales**

Cuando en el modelo de riesgos proporcionales se incluyen 2 o más factores, puede considerarse la conveniencia de incluir un término en el modelo que cuantifique el efecto para cada combinación de niveles de los factores. Estos efectos se conocen como interacciones. Por ejemplo, como se expone más adelante, supongamos que los dos factores son el nivel de  $L_p(a)$  (normal/patológico) y el nivel de LDL-

colesterol (bajo/alto), y que el centro del análisis es el tiempo hasta la reoclusión posterior a una operación de revascularización del sector aortofemoral en un grupo de pacientes. Si el efecto de la Lp(a) sobre el tiempo hasta la reoclusión depende del nivel (bajo o alto) del LDL-colesterol, se dice que hay interacción entre estos dos factores. La función de riesgo  $h(t)$  dependerá entonces de las combinaciones de niveles para estos dos factores.

Si los factores A y B tienen a y b niveles respectivamente, y si denotamos  $\alpha_j$  y  $\beta_k$  el efecto del j-ésimo nivel del factor A y el efecto del k-ésimo nivel del factor B, la interacción entre ambos la denotaremos  $(\alpha\beta)_{jk}$  para  $j = 1, 2, \dots, a$  y  $k = 1, 2, \dots, b$ , y esta interacción tendrá  $(a-1)(b-1)$  grados de libertad, es decir, número de parámetros asociados con la interacción. Obsérvese que, necesariamente, el término  $(\alpha\beta)_{jk} = 0$  si  $j=1$  ó  $k=1$ , siempre que hayamos considerado  $\alpha_1 = 0$  y  $\beta_1 = 0$ .

En general sólo se consideran modelos jerárquicos, es decir, modelos en los cuales se considera una interacción entre dos factores si cada uno de ellos por separado está también incluido en el modelo.

Los resultados de un análisis de datos de supervivencia utilizando el modelo de riesgos proporcionales, cuando se consideran interacciones entre factores, pueden presentarse de tres maneras distintas,<sup>3</sup> así, si los dos factores A y B tienen dos niveles cada uno, las tres opciones se presentan como se ilustra seguidamente:

Opción a) Aquí se hace evidente el efecto del factor B, para cada nivel del factor A.

factor A	factor B	
	nivel 1	nivel 2
nivel 1	1	$\exp(b_2)$
nivel 2	1	$\exp(b_2 + (ab)_{22})$

Opción b) Aquí se hace evidente el efecto del factor A, para cada nivel del factor B

factor A	factor B	
	nivel 1	nivel 2
nivel 1	1	1
nivel 2	$\exp(\alpha_2)$	$\exp(\alpha_2 + (\alpha\beta)_{22})$

Opción c) Aquí se hace evidente el efecto que tienen los dos factores cuando se toma como referencia la clase de sujetos que tienen el primer nivel en ambos factores.

factor A	factor B	
	nivel 1	nivel 2
nivel 1	1	$\exp(\beta_2)$
nivel 2	$\exp(\alpha_2)$	$\exp(\alpha_2 + \beta_2 + (\alpha\beta)_{22})$

### **Factores que afectan el tiempo hasta reoclusión de pacientes revascularizados**

Una muestra de 80 pacientes que fueron revascularizados del sector aortofemoral fueron estudiados prospectivamente para determinar el tiempo de supervivencia (en días) desde el momento de la operación hasta la eventual reoclusión (Trabajo de Terminación de Residencia, Lic. R. Simon y Dr. L. Chirino, Instituto Nacional de Angiología, 2000).

Al momento de la revascularización en cada paciente (tiempo cero) se midieron las siguientes variables:

- apolipoproteína A1 (APO-A1)
- apolipoproteína B (APO-B)
- colesterol total (CT)
- HDL colesterol (HDL-C)

- LDL colesterol (LDL-C)
- lipoproteína(a) (LP(A))
- triglicéridos (TRG)

El análisis de los datos se encontró en este caso en dos objetivos fundamentales:

- Determinar la influencia de LP(A) sobre el tiempo de supervivencia hasta la reclusión.
- Determinar la posible interacción entre la LP(A) y el LDL-C a los efectos de su influencia sobre el tiempo de supervivencia hasta la reclusión.

Se empleó el paquete estadístico SPSS for Windows v5,0 para todos los análisis estadísticos.

Para el primer objetivo se dicotomizó la LP(A) en “normal” y “patológico” según tomara valores menores que 30,6 y mayores o iguales a 30 respectivamente. Con posterioridad se aplicó la prueba *logrank* para comparar las funciones de supervivencia entre los dos grupos definidos anteriormente según el valor de LP(A). Los resultados se presentan en la tabla 1, los cuales son marginalmente significativos, por lo que hay cierta evidencia de que el tiempo de supervivencia depende del valor (dicotomizado en normal/patológico) de la LP(A).

Para el análisis de la interacción se utilizó la LP(A) dicotomizada como se indicó anteriormente; el LDL-C también fue dicotomizado en bajo y alto según el valor estuviera por debajo o no de la correspondiente mediana. Los resultados principales del ajuste del modelo de riesgos proporcionales con interacción entre estos dos factores se presentan en la tabla 2. Obsérvese que aunque el término correspondiente al factor LDL-C no es estadísticamente significativo ( $p = 0,46$ ), el principio de la modelación jerárquica nos obliga a mantener este factor en el modelo.

Los resultados anteriores nos permiten presentarlos de tres formas distintas, según se comentó con anterioridad (ver anexo).

De la opción b, en el ANEXO por ejemplo, podemos concluir que el efecto de la LPA sobre la función de riesgo (instantáneo de reclusión) es totalmente diferente según el valor del LDL-C. Para valores bajos de LDL-C, una LP(A) patológica aumenta el riesgo con respecto a una LP(A) normal en cerca de 13 veces, sin embargo, para valores altos de LDL-C una LP(A) patológica lejos de aumentar el riesgo lo disminuye en cerca de un tercio con respecto a una LP(A) normal.

Los resultados en el modelo sin interacción se presentan en la tabla 3.

**TABLA 1. Análisis de supervivencia para días desde la revascularización hasta la eventual reclusión**

	Total	Número de eventos	Número de censuras	% de censuras
LP(A) normal	35	3	32	91,43
LP(A) patológico	45	11	34	75,56
Total	80	14	66	82,50
Logrank	Estadígrafo 3,88	gl 1	p ,0488	

TABLA 2. Resultados del ajuste del modelo con interacciones

Variable	B	exp(B)	gl	p
LP(A)	2,5529	12,8	1	,0158
LDL-C	,8952	2,45	1	,4649
LDL-C* LP(A)	-2,9921	0,05	1	,0400

TABLA 3. Resultados del modelo sin interacción

Variable	B	exp(B)	gl	Sig
LP(A)	1,4549	4,3	1	,0275
LDL-C	-1,3514	0,26	1	,0247

Finalmente se observa que la no inclusión de la interacción modifica los resultados de varias maneras; las que a continuación se relacionan:

- Cada uno de los factores es estadísticamente significativo; en el modelo con interacción el factor LDL-C no se comportó así.
- El efecto “nocivo” de un varlo patológico de LP(A) con respecto a un valor normal es mucho menor en este momento que cuando se incluyó la interacción: 4,3 y 12,8 respectivamente.
- El efecto de un valor alto de LDL-C con respecto a un valor bajo es siempre protector independientemente del valor del otro factor LP(A); en el modelo con interacción el efecto de la LDL-C depende del valor de la LP(A).

Dado que la inclusión de la interacción en el modelo de riesgos proporcionales se presenta razonable, los resultados más plausibles son los correspondientes a este modelo por el contrario la no inclusión de la interacción en el modelo nos habría conducido a errar.

**SUMMARY:** The proportional hazards model (also known as Cox regression model) is presented for the analysis of survival data. The concept of interaction between 2 factors of this model is introduced and the options for presenting and interpreting the results are discussed. The approached procedures are illustrated by an example in which time is modelled until the reocclusion in aortofemoral revascularized patients, starting from some predictive variables, particularly lypoprotein(a) and LDL-cholesterol.

Subject headings: **PROPORTIONAL HAZARDS MODELS; SURVIVAL ANALYSIS; BIOMETRY.**

### Referencias bibliográficas

1. Cox DR. Regression models and life tables (with discusión). J Roy Stat Soc B 1972; 74:187-220.
2. Dobson AJ. Introduction to statistical modelling. London: Chapman and Hall; 1983: 24-25.
3. Collet D. Modelling survival data in medical research. London: Chapman and Hall; 1994:53-106.

Recibido: 17 de junio de 2002. Aprobado: 24 de julio de 2002.  
*Dr. Armando H. Seuc.* Instituto Nacional de Angiología y Cirugía Vascular. Calzada del Cerro No. 1551. Cerro. Ciudad de La Habana. Cuba. Teléf. 57-64-93.