



Investigación Clínica XXII. Del juicio clínico al modelo de riesgos proporcionales de Cox

Marcela Pérez-Rodríguez,^a Rodolfo Rivas-Ruiz,^a Lino Palacios-Cruz,^b Juan O. Talavera^a

Primero calcula, después, piensa...

SIR DAVID ROXBEE COX

Clinical research XXII. From clinical judgment to Cox proportional hazards model

Survival analyses are commonly used to determine the time of an event (for example, death). However, they can be used also for other clinical outcomes on the condition that these are dichotomous, for example healing time. These analyses only consider the relationship of one variable. However, Cox proportional hazards model is a multivariate analysis of the survival analysis, in which other potentially confounding covariates of the effect of the main maneuver studied, such as age, gender or disease stage, are taken into account. This analysis can include both quantitative and qualitative variables in the model. The measure of association used is called hazard ratio (HR) or relative risk ratio, which is not the same as the relative risk or odds ratio (OR). The difference is that the HR refers to the possibility that one of the groups develops the event before it is compared with the other group. The proportional hazards multivariate model of Cox is the most widely used in medicine when the phenomenon is studied in two dimensions: time and event.

Keywords Palabras clave

Survival	Supervivencia
Multivariate models	Modelos multivariados
Hazard ratio	Hazard ratio

Desde el artículo previo (capítulo XXI de esta serie), cuando hablábamos de curvas de supervivencia, mencionamos que en gran parte del ejercicio clínico relacionado con la atención médica la toma de decisiones implica el conocimiento del curso clínico de la enfermedad, es decir, la estimación del tiempo transcurrido hasta que un evento suceda.

Uno de los ejemplos más comunes es cuando un clínico desea estimar cuánto tiempo podría vivir un paciente con cáncer de pulmón con o sin tratamiento. Para realizar esta estimación se usa el método de Kaplan-Meier. Sin embargo, este método no permite plantear un modelo que tome en cuenta las distintas condiciones del paciente que pudieran modificar el desenlace, como la edad, la etapa clínica, etc. Además asume que los grupos son homogéneos con respecto a todas las variables; por ejemplo, asume que todos los sujetos tienen la misma edad, lo cual no es necesariamente el caso de los pacientes de los cuales queremos emitir un pronóstico confiable que nos permita realizar un juicio clínico adecuado.

A partir de este punto tenemos claro que, y lo hemos señalado previamente al tratar el tema del análisis de regresión logística múltiple (capítulo XX de esta serie), para que un modelo multivariado sea considerado completo debe ponderar el efecto de las distintas variables previamente consideradas, para la predicción de una o más variables dependientes, cuyo desenlace es dicotómico. Para este caso, el modelo de riesgos proporcionales de Cox (MRPC) constituye un modelo multivariado que puede ponderar el efecto de una serie de variables cualitativas o cuantitativas sobre un desenlace dicotómico a través del tiempo.

El MRPC fue creado y publicado en 1972, en la revista *Journal of the Royal Statistical Society* por el estadístico inglés Sir David R. Cox, como una alternativa al método de Kaplan-Meier en la que se incorporan las características del análisis de regresión en las tablas de supervivencia.

La función de riesgo para el MRPC puede expresarse mediante la siguiente ecuación:

$$\lambda(t; Z_i(t)) = \lambda_0(t) e^{\beta' Z_i(t)} \text{ ó } \lambda_1(t) = \lambda_0(t) e^{x_i(t) \beta}$$

^aCentro de Adiestramiento en Investigación Clínica (CAIC), Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social

^bSubdirección de Investigaciones Clínicas, Instituto Nacional de Psiquiatría "Dr. Ramón de la Fuente Muñiz", Secretaría de Salud

Distrito Federal, México

Comunicación con: Marcela Pérez-Rodríguez
Correo electrónico: marxelapr@gmail.com

Los análisis de supervivencia son usados comúnmente para establecer el tiempo de ocurrencia de un evento (por ejemplo, muerte). Sin embargo, pueden ser utilizados para otros desenlaces clínicos siempre y cuando estos sean dicotómicos, como tiempo de curación, tiempo de recaída, tiempo para que una enfermedad inicie, etc. Los análisis de Kaplan-Meier (K-M) solo consideran la relación de una variable a través del tiempo, mientras que los riesgos proporcionales de Cox son el modelo multivariado de este método, el cual toma en cuenta otras covariables posiblemente confusoras del efecto de la maniobra principal

estudiada, como la edad, el sexo o el estadio de la enfermedad. Este análisis puede incluir en su modelo variables dependientes cuantitativas y cualitativas. La medida de asociación que se usa se llama hazard ratio (HR) o razón de riesgos, la cual no es lo mismo que el riesgo relativo o la razón de momios (RM). La diferencia es que el HR se refiere a la posibilidad de que uno de los grupos llegue antes a un evento al compararlo con otro. El modelo de riesgos proporcionales de Cox es el modelo multivariado más usado en la medicina cuando se estudia el fenómeno en dos dimensiones: tiempo y evento.

Resumen

Donde $\lambda_0(t)$ es una función de riesgo no negativa y sin especificar, común a todos los sujetos del estudio, llamada *función de riesgo base*, y β es el vector de coeficientes del modelo. Es decir, el MRPC se utiliza fundamentalmente para comparar grupos y calcular cocientes de riesgo.

$$\text{Hazard} = \lambda = \frac{\text{Sucesos ocurridos en el instante } t}{\text{Sucesos en riesgo en el instante } t} = \frac{d}{n}$$

Dicho de una manera sencilla, la principal innovación del modelo es ofrecer un método de análisis de supervivencia sin necesidad de especificar o estimar una función de riesgo basal, mediante el desarrollo del concepto de *probabilidad marginal*, una función de probabilidad que solo depende de los coeficientes del modelo. A diferencia de la regresión logística, en este modelo se empieza con el 100 % de los sujetos en una condición específica; por ejemplo, vivo cuando se busca muerte, o sin evento adverso alguno cuando se inicia una terapia.

Como sabemos, cuando alguien muere no podemos atribuirlo a una sola causa, sino a múltiples factores como los que se encuentran al inicio de la enfermedad (o estado basal), los cuales pudieran corresponder a la demarcación pronóstica. Para ejemplificar lo que estamos comentando, pensemos en un caso hipotético en el cual estamos estudiando pacientes con cáncer de pulmón, y en los cuales queremos determinar qué variables se relacionan con el tiempo en que van a morir (variable dependiente o desenlace), tomando en cuenta por un lado que en el estado basal encontraríamos características como edad y sexo de los pacientes, estadio de la enfermedad y tabaquismo. En este caso, para poder estimar un desenlace con mayor certeza, podríamos tomar en cuenta estas variables para determinar el probable tiempo de supervivencia de un paciente, dependiendo de la presencia o ausencia de dichas características.

Al usar este modelo, el resultado se expresa por medio de una estadística llamada hazard ratio (HR), que es una división de riesgo (h) entre grupos:

$$\text{HR} = \frac{\text{Sucesos ocurridos en el instante } t \text{ en el grupo 1} / \text{Sujeto en riesgo en el instante } t \text{ en el grupo 1}}{\text{Sucesos ocurridos en el instante } t \text{ en el grupo 2} / \text{Sujeto en riesgo en el instante } t \text{ en el grupo 2}} = \frac{h1}{h2}$$

El HR, al igual que la razón de momios y el riesgo relativo, es una medida de riesgo; sin embargo, el HR se refiere a la probabilidad que tiene que presentar el desenlace en un periodo de tiempo en comparación con el grupo de referencia o control, es decir, es una probabilidad de tener el desenlace en un periodo de tiempo determinado. En los análisis de supervivencia, la mejor medida de tiempo que se debe considerar es la mediana de tiempo, en la que el 50 % de los participantes habrán desarrollado el desenlace.

Existen tres objetivos estadísticos que típicamente son considerados en el MRPC:

1. La prueba de significación de la maniobra que será la variable de agrupación, por ejemplo, el grupo de tratamiento, edad, estadio.
2. La obtención del estimador del efecto de la maniobra o variable de agrupación, que en este caso es el HR.
3. La obtención del intervalo de confianza para el HR.

Para probar la significación del efecto de la variable de agrupación o de tratamiento se utiliza la prueba de Wald, que proporciona un valor de p . Los supuestos que deben cumplirse en el modelo de Cox son los siguientes:

- Los datos censurados no están relacionados con la probabilidad de que ocurra el evento; es decir, en los sujetos cuyo desenlace desconocemos (sea cual sea la razón por la que se censuraron sus datos), se

presume que la falta de datos no se relaciona con el evento del desenlace.

- Las curvas de supervivencia para cada uno de los estratos (tratamientos u otra variable de agrupación) deben tener funciones de riesgo que sean proporcionales en el tiempo; esto debido a que es un modelo lineal y lo que significa es que cada cambio en el desenlace es proporcional a cada cambio en la variable de riesgo.

Una vez explicado lo anterior, regresemos al ejemplo del estudio en el cual pretendemos predecir el tiempo de muerte para cáncer de pulmón. Supongamos que tenemos una cohorte de 130 pacientes (hom-

bres y mujeres) con cáncer de pulmón en diferentes estadios de la enfermedad y queremos saber cuál es el riesgo de mortalidad en cada uno de esos estadios con base en un ajuste de diferentes datos clínicos que probablemente influyen en el desenlace (muerte), como la edad, el sexo y el tabaquismo.

La figura 1 muestra los pasos para realizar el análisis de riesgos proporcionales de Cox en el programa SPSS:

- El primer paso consiste en seleccionar del menú principal la opción Analizar.
- Sespués seleccionar Supervivencia y posteriormente Regresión de Cox.

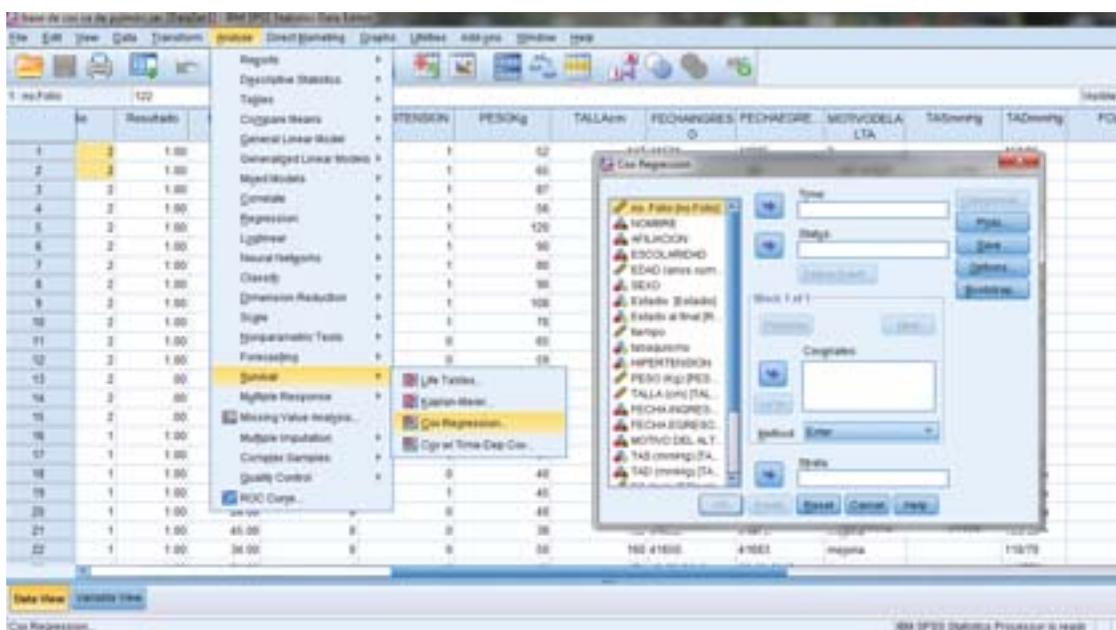


Figura 1 Pasos en el programa estadístico SPSS para realizar el análisis de riesgos proporcionales de Cox

		Frecuencia		
Sexo	1 = Mujer	62	1	
	2 = Hombre	68	0	
Estadio	0 = Etapa I	65	0	0
	1 = Etapa II	50	1	0
	2 = Etapa III	15	0	1
Tabaquismo	0 = ausente	77	0	
	1= presente	53	1	

- Variable de categoría: Sexo
- Codificación de parámetros de indicador
- Variabilidad de categoría: estadio (Estadio)
- Variable de categoría: tabaquismo

Figura 2 Codificación de las variables categóricas^{a,c,d}

En la figura 2 se presentan la frecuencia de las variables categóricas y la forma como se codificaron dichas variables para el análisis.

Una vez que está abierto el cuadro de diálogo (figura 3), el siguiente paso consiste en escoger las variables que queremos ingresar en el modelo. Para esto, lo primero que tenemos que hacer es agregar la variable tiempo, que en este ejemplo se refiere al número de meses que los pacientes con cáncer de pulmón permanecieron con vida. Posteriormente se tiene que definir la variable de resultado en la casilla de Estatus. Habitualmente el evento se codifica con el número 1 (presente). El siguiente paso es seleccionar las covariables que se van a introducir en el modelo, que en este caso son edad, sexo, estadio de la enfermedad y tabaquismo; de estas variables, las últimas tres

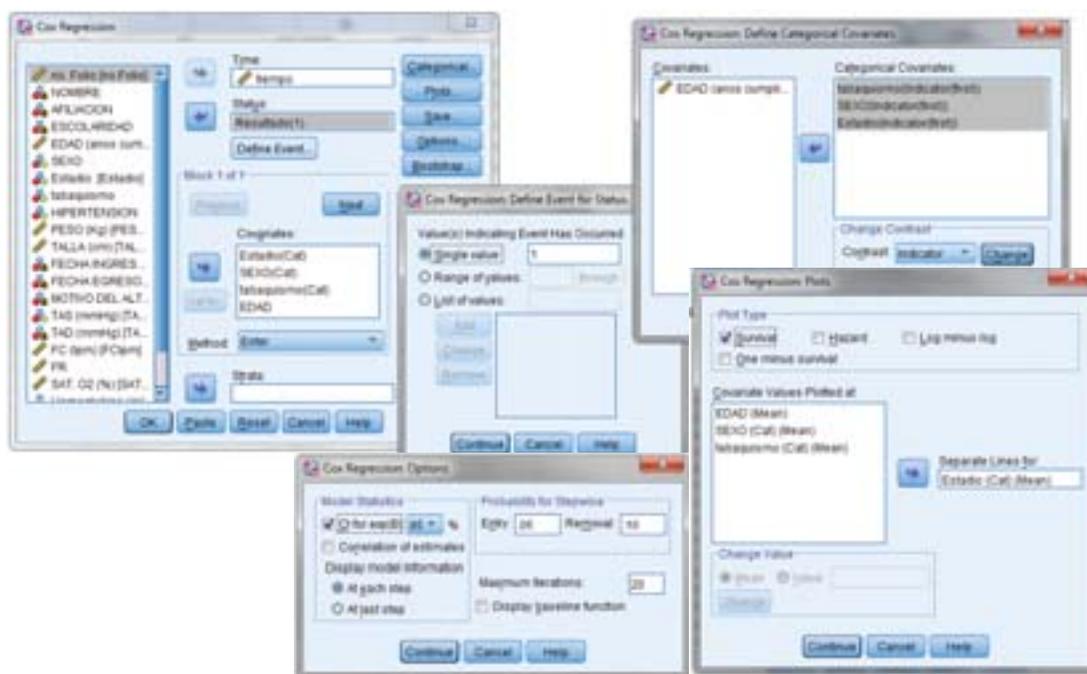


Figura 3 Pasos para la selección de variables categóricas y gráficos en el programa estadístico SPSS para realizar el análisis de riesgos proporcionales de Cox

son categóricas y, por lo tanto, hay que definir las en la opción de Categóricas y determinar cuál de las categorías de cada variable se va a tomar como referente al calcular los riesgos. En este caso nosotros tomamos la primera categoría de referencia, es decir, en la variable sexo las mujeres que están codificadas con el número 1 son el referente, el estadio 1 que está codificado con un 0 será el referente y tabaquismo ausente codificado con un 0 será la referencia. Una vez que ya está definido el evento y seleccionadas las variables que van a entrar al modelo, procedemos a seleccionar los gráficos. Para obtener la curva de supervivencia global se tiene que seleccionar la opción de Supervivencia. Para obtener además las curvas para cada uno

de los estadios de la enfermedad, podemos solicitar que se nos dé la curva con líneas separadas para la o las variables categóricas que deseemos. Esta curva nos dará el resultado de la supervivencia de ese grupo determinado, ajustado por las variables que componen el modelo. Por último, pero no por eso menos importante, debemos pedirle al programa que nos dé los intervalos de confianza al 95 % en el botón Opciones.

La figura 4 muestra la curva de supervivencia global en la que se representa gráficamente cómo van llegando al desenlace las personas conforme pasa el tiempo. Para el mes 60 de seguimiento menos del 40 % de la cohorte continúa con vida. Para la construcción de este gráfico el programa estadístico utiliza

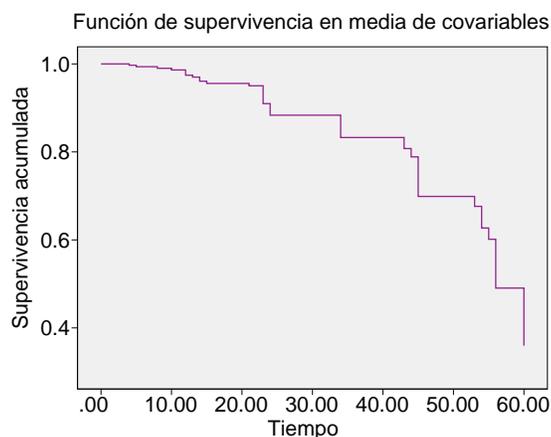


Figura 4 Curva de supervivencia global ajustada por los cofactores

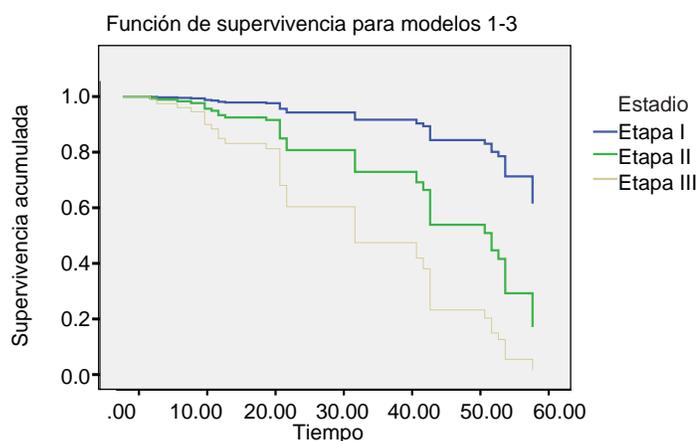


Figura 5 Curva de supervivencia por estadio de la enfermedad

las medias de edad, de estadio de la enfermedad, sexo y tabaquismo como covariables, es decir, no se presentan los riesgos para cada uno de los estadios de la enfermedad.

La figura 5 presenta las curvas de supervivencia ajustadas por edad, sexo y tabaquismo para cada uno de los estadios de la enfermedad, es decir, podemos observar cómo va incrementando la mortalidad en los pacientes en estadio I, II y III de la enfermedad. Es claro que los pacientes en estadio III viven menos que los pacientes en estadios menos avanzados. Para el mes 60 todos los pacientes en el estadio III han presentado el desenlace (muerte), mientras que cerca del 60 % de los pacientes con estadio I continúan vivos.

	Variable en la ecuación					
	B	Wald	Sig	Exp (B)	Inferior	Superior
Edad	.065	31.829	.000	1.067	1.043	1.091
Estadio I		15.843	.000			
Estadio II	1.289	7.230	.007	3.630	1.418	9.291
Estadio III	2.148	15.795	.000	8.568	2.970	24.712
Tabaquismo	.145	.136	.712	1.156	.536	2.492
Sexo (masc)	-.210	.459	.498	.810	.441	1.489

Figura 6 Resultados del modelo de riesgos proporcionales de Cox

La figura 6 muestra los resultados del análisis de Cox. Como en todos los modelos de regresión obtenemos valores de β . En este caso el exponente de β corresponde al HR. En este modelo, la edad y el estadio de la enfermedad son los principales factores de riesgo. La interpretación del HR sería de la siguiente manera: los pacientes en estadio II tienen 3.6 veces el riesgo (o 2.6 veces más riesgo) de haber

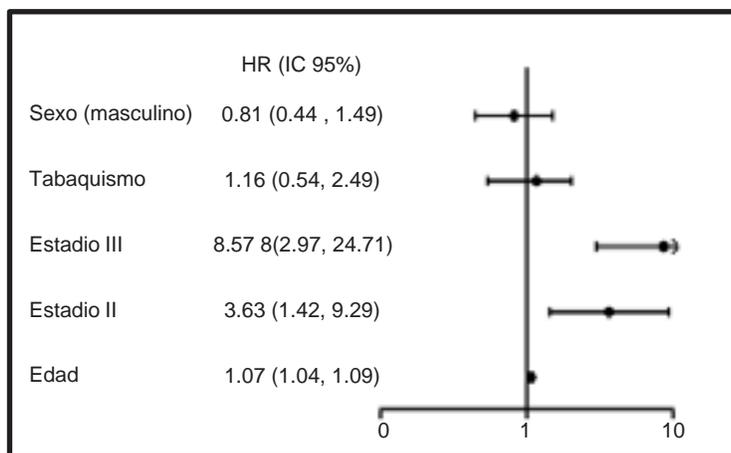
fallecido en el mes 60 de seguimiento en comparación con los que están en estadio I, y los pacientes en estadio III tienen 8.5 veces el riesgo (o 7.5 veces más riesgo) de haber fallecido para el mes 60 de seguimiento en comparación con el estadio I. Cada unidad de la variable edad (cada mes de vida) incrementa el riesgo de muerte en 6.7 %.

Estos HR también se pueden exponer de forma gráfica, usando *forest plots* para mostrar su relación con otras variables, donde la línea vertical en el número 1, es la no asociación (figura 7). Se interpreta de modo similar al OR o al RR. (capítulo VI de esta serie).

Este gráfico permite evaluar la significación estadística a partir de los intervalos de confianza (IC) de 95 %. Cuando el IC 95 % no toca la unidad (línea vertical) representa un estudio que es estadísticamente significativo, como el caso de la edad, el estadio II, y el estadio III. Sin embargo, cuando el IC 95 % toca la unidad e incluso la atraviesa, como en el caso de los pacientes fumadores o el ser hombres, estas variables no tienen relación con el desenlace o no contribuyen con el desenlace, que en este caso es la muerte.

Estos gráficos son muy populares, ya que permiten analizar de un vistazo la relación de las variables y su fuerza de asociación, es decir, no tan solo reflexionar sobre la significación estadística de la relación de las variables tomando en cuenta solo el valor de p , sino lo alejado de la no asociación (HR = 1), esto es, el tamaño del efecto, como lo es el estadio III de la enfermedad. Estos gráficos también permiten elaborar nuevas preguntas sobre el tamaño de muestra del estudio, ya que los intervalos de confianza dependen del número de pacientes incluidos, es decir en algunos casos el incrementar el número de observaciones (tamaño de muestra) puede disminuir la amplitud del intervalo de confianza. Dicho de otra manera, con estos gráficos se puede explicar la asociación y el tamaño del efecto de un solo vistazo.

Una vez finalizado el modelo de Cox, como se menciona en el figura 1 debe verificarse su validez y pertinencia en nuestros datos. El primer supuesto tiene que ver con los datos censurados. Debemos asegurarnos de que estos no están relacionados con la probabilidad de que ocurra el evento, es decir, que los sujetos que se pierden en el seguimiento no abandonaron el estudio por razones relacionadas con el tratamiento o la variable de agrupación, como puede ser el estadio de la enfermedad. Este supuesto se verifica investigando en la medida de lo posible las razones de abandono del estudio. Cuando este supuesto se viola, se invalidan todos los análisis de supervivencia desde el modelo de Kaplan-Meier hasta el modelo de Cox.



HR = hazard ratio; IC= intervalo de confianza

Figura 7 Gráfico de las covariantes que afectan la mortalidad por cáncer de pulmón

El segundo supuesto tiene que ver con que los riesgos entre los grupos sean proporcionales. Esto se aplica cuando la variable es cuantitativa y ordinal, y significa que las curvas de supervivencia para cada uno de los estratos deben tener funciones de riesgo que sean proporcionales en el tiempo; por ejemplo, que tengan riesgos con incrementos constantes de un valor a otro (de 50 a 60 años un incremento proporcional similar que de 60 a 70 años). Y que siga la misma dirección, es decir a mayor estadio de la enfermedad, mayor riesgo. Esto puede evaluarse visualmente con un gráfico de log-log (Figura 8) que se solicita en la opción de gráficos en el cuadro de diálogo del modelo de Cox. En nuestro ejemplo, este gráfico muestra que la función de riesgos sí es proporcional entre las etapas de la enfermedad.

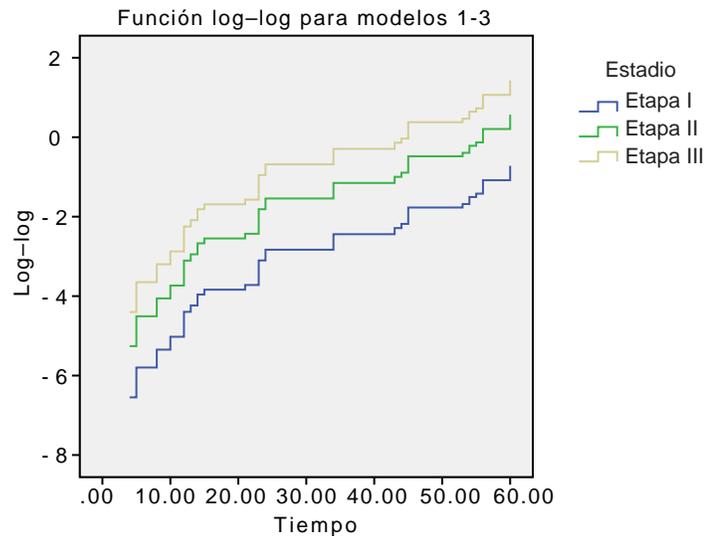


Figura 8 Gráfico de log-log para evaluar función de riesgos proporcionales

Conclusiones

Los modelos de supervivencia son alternativas estadísticas que nos permiten hacer la estimación de un riesgo tomando en cuenta el tiempo transcurrido hasta el desenlace. Este tipo de modelos son bastante útiles en el campo de la salud, especialmente en áreas relacionadas con enfermedades crónicas y enfermedades de alta letalidad. El modelo de riesgos proporcionales de Cox constituye un modelo multivariado que puede ponderar una curva de supervivencia con variables cualitativas o cuantitativas que pueden modificar el efecto final.

Tomar en cuenta todos los posibles factores de riesgo que puedan intervenir en un resultado ayuda a fundamentar el juicio clínico.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no ha sido reportado alguno que esté relacionado con este artículo.

Referencias

1. Kleinbaum DG, Klein M. *Survival Analysis: A Self-Learning Text*. New York: Springer Science Business Media, Inc; 2005.
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-81.
3. Feinstein AR. *Principles of medical statistics*. New York, NY: Chapman and Hall/CRC; 2002.
4. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: Multivariate data analysis – an introduction to concepts and methods. *Br J Cancer*. 2003;89:431-6.
5. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972;34(2):187-220.
6. Berea-Baltierra R, Rivas-Ruiz R, Pérez-Rodríguez M, Palacios-Cruz L, Moreno J, Talavera JO. Del juicio clínico a la regresión logística múltiple. *Rev Med Inst Mex Seguro Soc*. 2014;52(2):192-7.