



## FORMACIÓN CONTINUADA - METODOLOGÍA Y TÉCNICAS

# Análisis de supervivencia

J.A. Martínez Pérez<sup>a,\*</sup> y P.S. Pérez Martínez<sup>b</sup>



<sup>a</sup> Comisión Nacional de Calidad de Semergen, España

<sup>b</sup> Centro de Salud Guzmán el Bueno, Madrid, España

Recibido el 17 de enero de 2023; aceptado el 3 de abril de 2023

Disponible en Internet el 12 de mayo de 2023

### PALABRAS CLAVE

Evento inicial y final;  
Tiempos incompletos;  
Kaplan-Meier;  
Método actuarial

**Resumen** El análisis de supervivencia es un método estadístico que valora el tiempo entre un evento inicial (inclusión del sujeto en el estudio) y uno final, que sucede cuando este presenta una característica definida con anterioridad (evento). Su objetivo es estimar, teniendo en cuenta la variable tiempo, la probabilidad de que ocurra un suceso determinado. Tiene la particularidad de aceptar tiempos incompletos de participación y asumir que todos los factores implicados en el estudio son homogéneos. Existen varios métodos para calcular la probabilidad de supervivencia; los más utilizados son los de Kaplan-Meier y el actuarial.

© 2023 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

### KEYWORDS

Initial and final event;  
Incomplete times;  
Kaplan-Meier;  
Actuarial method

### Survival analysis

**Abstract** Survival analysis is a statistical method that assesses the time between an initial event (inclusion of the subject in the study) and a final event, which occurs when the subject presents a previously defined characteristic. Its objective is to estimate, taking into account the time variable, the probability of a certain event occurring. It has the particularity of accepting incomplete participation times and assuming that all the factors involved in the study are homogeneous. There are several methods to calculate the probability of survival, the most used are the Kaplan-Meier and the actuarial.

© 2023 Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Published by Elsevier España, S.L.U. All rights reserved.

En medicina se suele aplicar el análisis de supervivencia con el fin de valorar la supervivencia de los pacientes a una determinada enfermedad. Este tipo de estudios pertenecen a los llamados longitudinales, en los que los individuos se siguen a lo largo del tiempo; en este caso, el seguimiento

\* Autor para correspondencia.

Correo electrónico: [jmartinezp@semergen.es](mailto:jmartinezp@semergen.es)  
(J.A. Martínez Pérez).

se hace desde la entrada al estudio de un individuo hasta la ocurrencia de un evento de interés<sup>1</sup>.

Este evento no es necesariamente la muerte, puede ser tiempo de recurrencia, tiempo que dura la eficacia de una intervención, tiempo de remisión, tiempo de persistencia de un tratamiento<sup>2</sup>, tiempo de aparición de nuevos factores de riesgo en un seguimiento<sup>3</sup>, etc., lo que hace de estos estudios una herramienta muy útil en Atención Primaria. El vocablo supervivencia se debe a que en las primeras aplicaciones de este método de análisis se utilizaba como evento la muerte de un paciente.

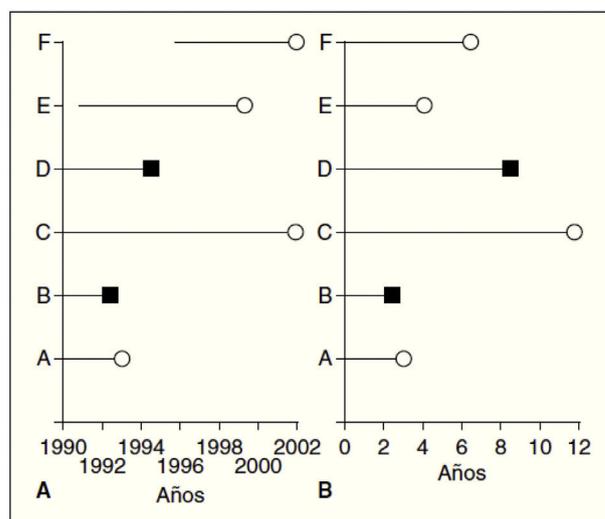
Estos estudios están especialmente diseñados para determinar la probabilidad de un suceso en diferentes intervalos de tiempo, siendo siempre el intervalo de tiempo la variable dependiente. Además, en él pueden participar pacientes con tiempos incompletos que no han llegado al estadio terminal<sup>1</sup>.

## Conceptos básicos

- Fecha de inicio.** Es aquella en que después de haber terminado los preparativos (protocolos, permisos, etc.), se empieza la inclusión de pacientes. Es igual para todos los pacientes.
- Fecha de entrada.** Cuando se incorpora al paciente en el estudio. Es muy importante porque a partir de ella se calcula el tiempo de supervivencia. Es diferente para cada sujeto.
- Fecha de última observación.** Última visita realizada. Tanto las fechas de entrada como de última observación son diferentes para cada paciente, puesto que estos se incorporan en diferentes momentos.
- Tiempo de participación.** El tiempo transcurrido entre la fecha de entrada y la fecha de última observación.
- Fecha de cierre.** Cuando se decide finalizar el estudio. Debe ser la misma para todos los pacientes.
- Duración del estudio.** El intervalo de tiempo discurrido entre la fecha de entrada y la fecha de cierre.
- Tiempo de supervivencia.** Se define como el intervalo de tiempo entre la entrada del paciente en el estudio y el evento de interés<sup>4</sup>.
- Paciente censurado.** Cuando su tiempo de participación termina antes de la fecha de cierre y por consiguiente su seguimiento es incompleto o perdido. El tiempo de participación puede terminar por los siguientes motivos:
  - El paciente decide no participar más en el estudio y lo abandona.
  - El paciente se pierde (por ejemplo, cambia de domicilio) y no tenemos más información sobre él.
  - El estudio termina antes de aparecer el evento que estudiamos<sup>5</sup> (fig. 1).

## Consideraciones importantes

Es necesario usar una definición operativa del punto de partida (operación quirúrgica, ingreso en el hospital, inicio de un tratamiento, etc.) para fijar la fecha de entrada en el estudio. No se deben usar situaciones ambiguas (por ejemplo, comienzo de sintomatología) que dependan de la memoria del paciente. Debe ser una fecha exacta.



**Figura 1** Esquema del tiempo de supervivencia tras la entrada en un estudio. Tomado de Abaira<sup>1</sup>.

El estudio empezó el 1 de enero de 1990 y acabó el 1 de enero de 2002. En A el eje temporal representa años de calendario y en B años desde el diagnóstico. Con el círculo en blanco se representan los tiempos censurados y con el cuadrado negro las muertes (ocurrencia del evento). El individuo A al que se le diagnosticó en enero de 1990 desapareció del estudio en enero de 1993 (sería una censura a los 3 años por pérdida de seguimiento). El B, también diagnosticado en enero de 1990, falleció en junio de 1992 (muerte a los 2,5 años). El C seguía vivo al acabar el estudio (sería un dato censurado a los 12 años por fin del estudio). El D, al que se le diagnosticó en febrero de 1991, falleció en marzo de 1999; el tiempo de supervivencia sería de 8 años. El E, que fue diagnosticado en noviembre de 1993, falleció en accidente de tráfico en julio de 1997 (sería una muerte, o un dato censurado, a los 3,7 dependiendo de la definición de evento de interés). El F, al que se le diagnosticó al principio de 1996, seguía vivo al acabar el estudio (sería un dato censurado a los 6 años por fin del estudio).

El evento de interés debe ser claro y muy bien definido para poder precisar la fecha del mismo. En caso de ser la muerte del paciente, es imprescindible que esta sea como consecuencia de la enfermedad objeto del estudio. Si no es así, el fallecido debe valorarse como censurado y su tiempo de seguimiento como incompleto. En caso contrario estaremos introduciendo en el estudio un sesgo de información.

En la última observación es preciso registrar el estado del sujeto y la fecha de información de dicho estado. Si el evento de interés es la muerte y el paciente ha fallecido, calcularemos el tiempo de supervivencia en base a la fecha de defunción. Si el paciente está vivo a la fecha de la última observación, se calculará como tiempo incompleto o censurado, porque lo que define un tiempo completo es la presencia del evento de estudio.

De esta manera, podemos decir que los requisitos imprescindibles para disponer de datos idóneos para un análisis de supervivencia son<sup>5</sup>:

1. Definir adecuadamente el origen o inicio del seguimiento.

2. Definir adecuadamente la escala del tiempo.
3. Definir adecuadamente el evento de interés.

## Sesgos. Tipos de observaciones

Los datos del estudio pueden estar sesgados por las censuras y los truncamientos. Como hemos visto anteriormente, en las censuras el tiempo de seguimiento termina antes de producirse el evento de interés o cuando este no se produjo.

En los truncamientos, se entró al estudio después del hecho que define el origen en todos los individuos analizados. Se tendría que haber empezado con anterioridad ya que la enfermedad habría comenzado antes.

De esta forma, se pueden producir cuatro tipos de observaciones de datos<sup>6</sup>:

1. No truncada, no censurada: el proceso se inicia en el inicio del seguimiento y el evento aparece durante el seguimiento realizado.
2. No truncada, censurada: el proceso se inicia en el inicio del seguimiento, pero el evento no aparece durante el seguimiento realizado.
3. Truncada no censurada: el paciente ya tenía el proceso cuando entró al estudio (lo hizo tarde) y el evento aparece durante el seguimiento realizado.
4. Truncada, censurada: el paciente ya tenía el proceso cuando entró al estudio y el evento no aparece durante el seguimiento realizado.

## Supervivencia y riesgo

Los datos de supervivencia se pueden dar y estudiar con dos tipos de probabilidades diferentes: supervivencia y riesgo. La «probabilidad de supervivencia» (también llamada función de supervivencia) o  $S(t)$  es la probabilidad de que un individuo sobreviva desde la fecha de entrada en el estudio hasta un momento determinado en el tiempo  $t$ . Estos valores van a describir la supervivencia global de toda nuestra población. Esta función se centra sobre todo en la «no ocurrencia» del evento (el paciente no ha fallecido; o no ha recidivado) y se utiliza para encontrar la mediana (percentil 50) y otros percentiles de tiempo de vida, así como para comparar datos de supervivencia de dos o más grupos.

Pero más útil que la probabilidad de supervivencia es la «función de riesgo»  $h(t)$ : que es la probabilidad de que a un individuo que está siendo observado en el tiempo  $t$  le suceda el evento en ese momento. Es decir: ¿cuál es la probabilidad de que fallezca un paciente con cáncer de pulmón microcítico a los 12 meses (suponiendo que haya sobrevivido hasta ese momento)? La función de riesgo se centra en la «ocurrencia» del evento y proporciona información como la tasa de incidencia, por ejemplo: ¿en qué momento voy a tener el pico de recidivas?<sup>7</sup>.

## Análisis de datos

El análisis de datos para este tipo de estudios requiere métodos de análisis específicos debido a que frecuentemente se analizan los datos antes de que ocurra el evento terminal (si no el estudio duraría mucho tiempo), y por eso muchos de los

datos que se poseen son censurados y además los pacientes no entran en el estudio al mismo tiempo.

Hay que tener en cuenta que en la mayor parte de los estudios la variable tiempo no tiene una distribución normal, más bien suele tener una distribución asimétrica, en la que típicamente tenemos más datos al principio que al final del período de seguimiento<sup>1,8</sup>. En este caso suele utilizarse la prueba de Kaplan-Meier, aunque también existen otras pruebas no paramétricas como la de Log rango y la regresión de Cox<sup>5,8</sup>.

Cuando la variable tiempo sigue una distribución normal, se pueden utilizar las pruebas de distribución exponencial, distribución de Weibull o distribución Log normal<sup>5</sup>.

Indudablemente los métodos estadísticos más utilizados en el análisis de supervivencia son los de Kaplan-Meier y el actuarial<sup>9</sup>.

## Método de Kaplan-Meier

Conocido también como del «límite del producto». En él suponemos que los sujetos censurados se habrían comportado del mismo modo que los seguidos hasta que se produjo el evento y asumimos que este evento terminal es independiente para cada paciente, lo que hace que las probabilidades de sobrevivir en un tiempo determinado se calculan gracias a una ley multiplicativa de probabilidades.

Aunque la supervivencia en un tiempo dado depende de la supervivencia en todos los períodos previos, la posibilidad de la misma en un período de tiempo es independiente de la probabilidad de supervivencia en los demás períodos. Otra característica distintiva es que no se agrupan los tiempos de supervivencia en intervalos. Debido a ello, este método es especialmente útil en estudios con un número pequeño de pacientes<sup>5,6</sup>.

## Método actuarial de Cutler-Ederer

Los tiempos de supervivencia se agrupan en intervalos. La longitud del intervalo depende de la frecuencia con que ocurre el evento de interés y puede no ser necesariamente de la misma longitud. La desventaja principal de este método es que da estimaciones poco precisas cuando el número de sujetos es pequeño. En muestras grandes, tiene una importancia fundamental porque influye poco sobre las estimaciones y además permite obtener una función de riesgo. Se asume que las retiradas y pérdidas se distribuyen homogéneamente en el intervalo, pero puede establecer un sesgo muy importante cuando los intervalos son muy grandes o hay muchos abandonos<sup>6</sup>.

## Cálculo de las curvas de supervivencia en SPSS

### Método de Kaplan-Meier

En un estudio de 62 pacientes con cáncer de pulmón, hemos estimado el tiempo de supervivencia de cada uno de ellos y hemos anotado su estado al final del seguimiento (el evento de interés era el fallecimiento) y el tipo celular del tumor que tenían. Los datos de estas variables los hemos codificado de la siguiente manera:

**Resumen de procesamiento de casos**

N total	N de eventos	Censurado	
		N	Porcentaje
62	57	5	8,1%

**Tabla de supervivencia**

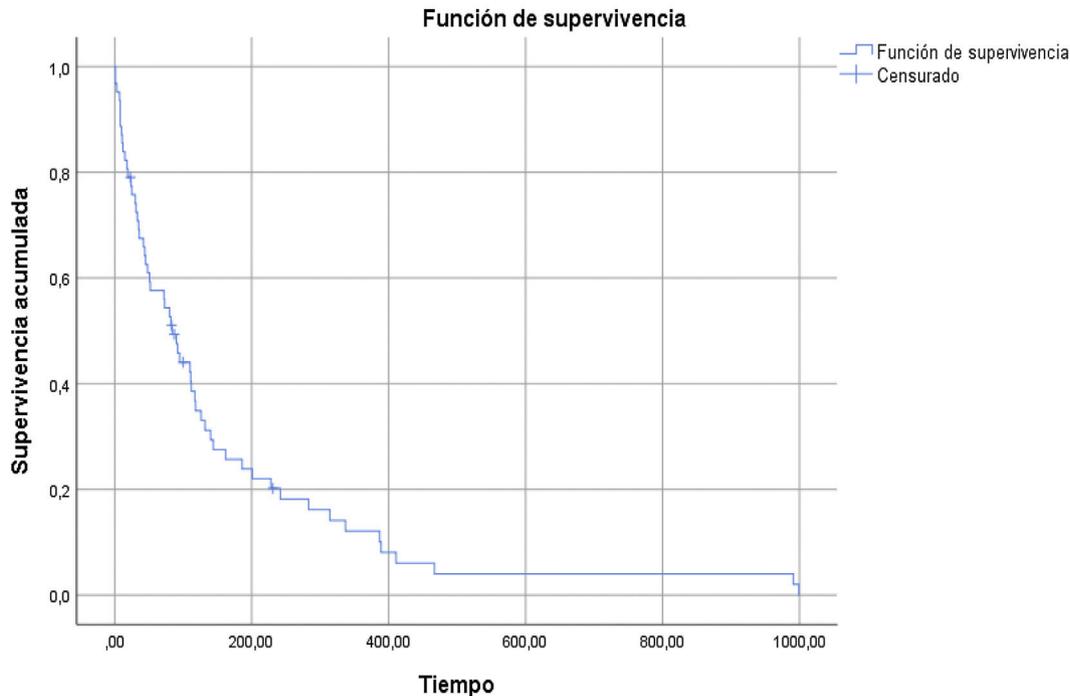
	Hora	Estado	Proporción acumulada que sobrevive en el tiempo		N de eventos acumulados	N de casos restantes
			Estimación	Desv. Error		
1	1,000	Fallecido	.	.	1	61
2	1,000	Fallecido	,968	,022	2	60
3	3,000	Fallecido	,952	,027	3	59
4	7,000	Fallecido	,935	,031	4	58
5	8,000	Fallecido	.	.	5	57
6	8,000	Fallecido	.	.	6	56
7	8,000	Fallecido	,887	,040	7	55
8	10,000	Fallecido	,871	,043	8	54
9	11,000	Fallecido	,855	,045	9	53
10	12,000	Fallecido	,839	,047	10	52
11	15,000	Fallecido	,823	,049	11	51
12	18,000	Fallecido	,806	,050	12	50
13	19,000	Fallecido	,790	,052	13	49
14	23,000	Censurado	.	.	13	48
15	24,000	Fallecido	,774	,053	14	47
16	25,000	Fallecido	,757	,055	15	46
17	30,000	Fallecido	,741	,056	16	45
18	31,000	Fallecido	,724	,057	17	44
19	33,000	Fallecido	,708	,058	18	43
20	35,000	Fallecido	,692	,059	19	42
21	36,000	Fallecido	,675	,060	20	41
22	42,000	Fallecido	,659	,061	21	40
23	44,000	Fallecido	,642	,061	22	39
24	45,000	Fallecido	,626	,062	23	38
25	48,000	Fallecido	,609	,062	24	37
26	51,000	Fallecido	,593	,063	25	36
27	52,000	Fallecido	,576	,063	26	35
28	72,000	Fallecido	,560	,063	27	34
29	73,000	Fallecido	,543	,064	28	33
30	80,000	Fallecido	,527	,064	29	32
31	82,000	Fallecido	,510	,064	30	31
32	83,000	Censurado	.	.	30	30
33	84,000	Fallecido	,493	,064	31	29
34	87,000	Censurado	.	.	31	28
35	90,000	Fallecido	,476	,064	32	27
36	92,000	Fallecido	,458	,064	33	26
37	95,000	Fallecido	,441	,064	34	25
38	100,000	Censurado	.	.	34	24
39	110,000	Fallecido	,422	,064	35	23
40	111,000	Fallecido	,404	,064	36	22
41	112,000	Fallecido	,385	,063	37	21
42	117,000	Fallecido	,367	,063	38	20
43	118,000	Fallecido	,349	,062	39	19
44	126,000	Fallecido	,330	,062	40	18
45	132,000	Fallecido	,312	,061	41	17
46	140,000	Fallecido	,294	,060	42	16
47	144,000	Fallecido	,275	,059	43	15
48	162,000	Fallecido	,257	,058	44	14
49	186,000	Fallecido	,239	,057	45	13
50	201,000	Fallecido	,220	,055	46	12
51	228,000	Fallecido	,202	,054	47	11
52	231,000	Censurado	.	.	47	10
53	242,000	Fallecido	,182	,052	48	9
54	283,000	Fallecido	,162	,050	49	8
55	314,000	Fallecido	,141	,048	50	7
56	337,000	Fallecido	,121	,045	51	6
57	387,000	Fallecido	,101	,042	52	5
58	389,000	Fallecido	,081	,038	53	4
59	411,000	Fallecido	,061	,033	54	3
60	467,000	Fallecido	,040	,028	55	2
61	991,000	Fallecido	,020	,020	56	1
62	999,000	Fallecido	,000	,000	57	0

Figura 2 Resumen de procesamiento de casos y tabla de supervivencia.

**Medias y medianas para el tiempo de supervivencia**

Media <sup>a</sup>				Mediana			
Estimación	Desv. Error	Intervalo de confianza de 95 %		Estimación	Desv. Error	Intervalo de confianza de 95 %	
		Límite inferior	Límite superior			Límite inferior	Límite superior
153,443	28,589	97,409	209,477	84,000	12,352	59,791	108,209

a. La estimación está limitada al tiempo de supervivencia más largo, si está censurado.



**Figura 3** Medias y medianas del tiempo de supervivencia y función de supervivencia.

Tiempo de supervivencia en días  
 Estado - 1: fallecido; 0: censurado  
 Tipo de célula - 1: escamoso; 2: adenocarcinoma

**Ejecución del programa**

Analizar → Supervivencia → Kaplan-Meier. Posteriormente aparecerá una nueva ventana donde introduciremos, en la casilla «Hora», la variable Tiempo, y en la casilla «Estado», como su propio nombre indica, la variable estado en donde hemos registrado si se presentó el suceso de interés o fue un dato censurado.

Definir evento → Valor único (ponemos el código que le hemos asignado al evento de interés, que en nuestro caso es 1) → Continuar

Opciones → marcamos tablas de supervivencia, media y mediana de supervivencia y en gráficos: supervivencia, después Continuar → Aceptar

De este modo obtendremos (figs. 2 y 3):

- Un resumen del procesamiento de casos: número total de intervinientes en el estudio, número de eventos de interés, número de censurados y porcentaje de estos últimos.
- La tabla de supervivencia ordenada de menor a mayor por tiempos de supervivencia: hora (en este caso tiempo

en días), estado, proporción acumulada de personas que sobreviven en el tiempo con su desviación típica, número de eventos acumulados y número de casos restantes.

- Medias y medianas para el tiempo de supervivencia: estimación de la media y mediana con su desviación típica e intervalo de confianza al 95%.
- Gráfico de la curva de supervivencia.

**Interpretación**

Si observamos las columnas, hora (tiempo) y estimación de la tabla (fig. 2), podemos determinar la probabilidad de que una persona viva según el intervalo de tiempo que queramos averiguar. Así, podemos decir que en el primer mes (30 días) existe una probabilidad del 74,1% de supervivencia, del 47,6% a los 3 meses (90 días) y de algo más de un 10,1% al año.

Por otra parte, al tratarse de una distribución paramétrica de datos es preferible utilizar la mediana (percentil 50) a la media, para precisar cuál es el tiempo de supervivencia del 50% de la población estudiada. La media es una medida que describe la tendencia central de una distribución, sin embargo, en este caso es más adecuado usar la mediana por la menor influencia que ejercen en ella los valores extremos de la variable tiempo.

**Medias y medianas para el tiempo de supervivencia**

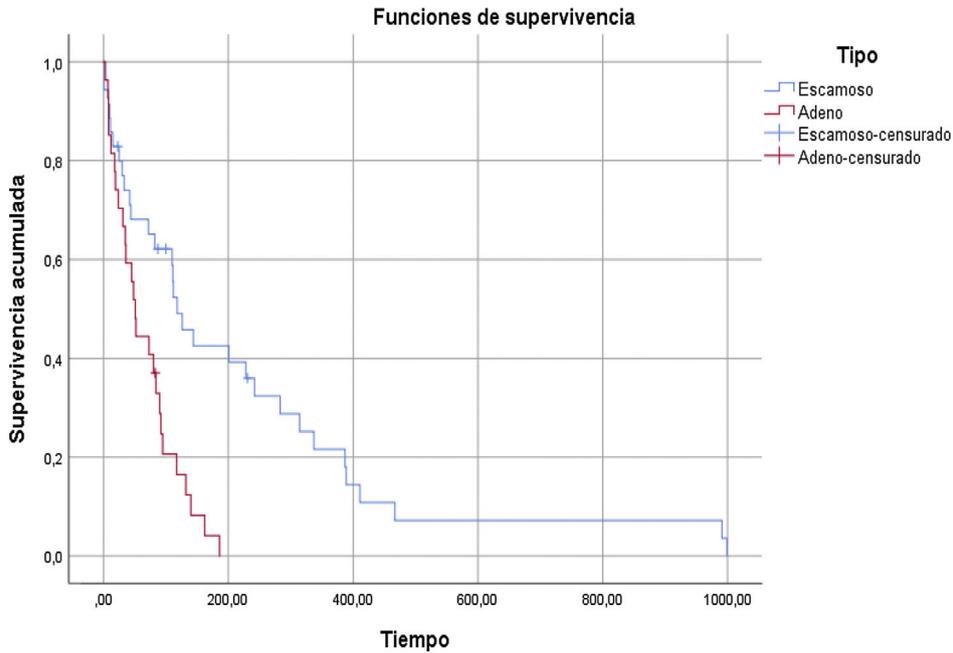
Tipo	Estimación	Desv. Error	Media <sup>a</sup>		Estimación	Desv. Error	Mediana	
			Intervalo de confianza de 95 %				Intervalo de confianza de 95 %	
			Límite inferior	Límite superior			Límite inferior	Límite superior
Escamoso	222,044	47,148	129,634	314,455	118,000	22,086	74,712	161,288
Adeno	65,556	10,127	45,707	85,404	51,000	6,058	39,126	62,874
Global	153,443	28,589	97,409	209,477	84,000	12,352	59,791	108,209

a. La estimación está limitada al tiempo de supervivencia más largo, si está censurado.

**Comparaciones globales**

	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	12,004	1	,001

Prueba de igualdad de distribuciones de supervivencia para los distintos niveles de Tipo.



**Figura 4** Comparación entre dos curvas de supervivencia.

En este estudio el 50% de la población tenía un tiempo de supervivencia de 84 días (fig. 3). Respecto a la curva, esta representa un tiempo de supervivencia corto.

**Comparación de dos curvas**

Si en nuestro estudio además queremos saber en cuál de los dos tumores (escamoso y adenocarcinoma) es mayor el tiempo de supervivencia, tenemos que comparar las curvas de supervivencia de cada uno de ellos.

Para ello seguimos el mismo procedimiento que en el apartado anterior, es decir: Analizar → Supervivencia → Kaplan-Meier, a continuación introduciremos en la casilla «Hora» la variable Tiempo y en la casilla «Estado», la

variable estado. Después: Definir evento → Valor único (ponemos el código que le hemos asignado al evento de interés) → Continuar

Pero en esta ocasión: Factor (introducimos la variable por la cual queremos saber si hay diferencia entre las curvas de supervivencia, que en nuestro caso es Tipo) → Comparar factor → Log rango → Continuar

Opciones → marcamos tablas de supervivencia, media y mediana de supervivencia y en gráficos: supervivencia, después Continuar → Aceptar

De esta manera obtendremos (fig. 4):

- Medias y medianas para el tiempo de supervivencia de ambos tumores con sus desviaciones típicas e intervalos de confianza al 95%.

- Prueba de Log rango (Mantel-Cox) para determinar si hay diferencias significativas entre las dos variables.
- Gráfico de las dos curvas de supervivencia.

### Interpretación

Al analizar los resultados, observamos que hay una gran diferencia en lo que se refiere a tiempos de supervivencia entre la media de los tumores escamosos y los adenocarcinomas a favor de los primeros. Al aplicar el estadístico Log rango (Mantel-Cox), apreciamos que esta diferencia entre las dos curvas de supervivencia es estadísticamente muy significativa ( $p < 0,001$ ), por lo cual tenemos que rechazar la hipótesis nula que establecía que no había diferencias entre ambas, lo que significa que los pacientes con tumores de tipo escamoso tienen mayor esperanza de vida que los que tienen adenocarcinoma.

### Bibliografía

1. Abraira V. Análisis del tiempo hasta un evento (supervivencia). *Semerger*. 2004;30:223-5.
2. Sicras-Mainar A, Sicras-Navarro A. Persistencia al tratamiento con metformina de marca vs. genérica en monoterapia para la diabetes tipo 2: estudio retrospectivo de vida real mediante propensity score matching. *Semerger*. 2021;47:321-31.
3. Cinza Sanjurjo S, Llisterri Caro JL, Barquilla García A, Polo García J, Velilla Zancada S, Rodríguez Roca GC, et al. Descripción de la muestra, diseño y métodos del estudio para la identificación de la población española de riesgo cardiovascular y renal (IBERICAN). *Semerger*. 2020;46:4-15.
4. Schober P, Vetter TR. Survival Analysis and Interpretation of Time-to-Event Data: The Tortoise and the Hare. *Anesth Analg*. 2018;127:792-8.
5. Pita Fernandez S. Análisis de supervivencia. *Cad Aten Primaria*. 1995;2:130-5.
6. Arribalzaga EB. Interpretación de las curvas de supervivencia. *Rev Chil Cir*. 2007;59:75-83, <http://dx.doi.org/10.4067/S0718-40262007000100013>.
7. Rebaso P. Conceptos básicos del análisis de supervivencia. *Cir Esp*. 2005;78:222-3.
8. Indrayan A, Tripathi CB. Survival Analysis: Where, Why What and How? *Indian Pediatr*. 2022;59:74-9.
9. Barakat A, Aaina Mittal A, Ricketts D, Rogers BA. Understanding survival analysis: actuarial life tables and the Kaplan-Meier plot. *Br J Hosp Med (Lond)*. 2019;80:642-6.