

Introducción a los estudios de cohorte en epidemiología y al análisis de supervivencia

Francisco Javier Díaz Ceballos, PhD*

Resumen

Los diseños de cohorte se han constituido en una de las herramientas metodológicas más usadas en la investigación epidemiológica y médica. En este artículo, se hace una introducción elemental a los conceptos y los métodos estadísticos usados en estos diseños. Después de describir las ideas fundamentales de los estudios de cohorte prospectivos y retrospectivos, y de los estudios de cohorte fija y variable, el artículo se enfoca en las herramientas estadísticas más elementales usadas en los estudios de cohorte variables. Estas herramientas hacen parte de lo que se conoce como “análisis de supervivencia”. Se explica entonces el concepto de censura y se describen las tablas de vida, las funciones de supervivencia, el estimador de Kaplan-Meier y la función de riesgo acumulado. También se explican métodos estadísticos para comparar dos funciones de supervivencia, tales como la prueba de Cochran-Mantel-Haenszel, la prueba log-rank y la prueba log-rank estratificada. Finalmente, se dan los conceptos básicos del método alternativo conocido como “método persona-años”. Los conceptos y métodos se ilustran usando un estudio de cohorte retrospectivo de la edad de comienzo de fumar diariamente en pacientes esquizofrénicos. [Díaz FJ. *Introducción a los estudios de cohorte en epidemiología y al análisis de supervivencia*. MedUNAB 2005; 8(1):43-53].

Palabras clave: Diseño de cohorte, análisis de supervivencia, curvas de supervivencia, estimador de Kaplan-Meier, función de riesgo acumulado, prueba de Cochran-Mantel-Haenszel, prueba log-rank, método persona-años, fumar, esquizofrenia.

Introducción

Un estudio de cohorte consiste en hacerle un seguimiento a grupos de personas para monitorear su salud. El diseño más simple consiste en seguir dos grupos (llamados cohortes), uno de los cuales está expuesto a un factor de riesgo mientras que el otro no lo está, aunque idealmente los dos grupos deben tener características similares. Después de determinado período de tiempo, llamado ‘período de seguimiento’ (*follow-up period*), la incidencia de cierta enfermedad se compara entre los dos grupos. El comienzo del estudio se llama “tiempo de referencia” (*baseline*). Por ejemplo, el primer grupo puede ser una muestra aleatoria de empleados de una fábrica que utiliza asbestos, y el otro grupo una muestra aleatoria de empleados de otras fábricas similares pero que no utilizan asbestos. Durante cierto período de tiempo, digamos 10 años, se registran los casos de cáncer del pulmón en ambos grupos. Al finalizar este período de tiempo se comparan las incidencias de cáncer del pulmón en los dos grupos. Idealmente, ninguno de los individuos en los dos grupos debe mostrar síntomas de cáncer al comienzo del estudio, con el objetivo de poder establecer claramente relaciones de causalidad entre el factor de riesgo (exposición al asbestos) y la enfermedad. Además, el tipo de información recolectado durante el período de seguimiento puede ser el número de casos de enfermedad observados (morbilidad), o el número de muertes debidas a la enfermedad (mortalidad).

Un diseño alternativo para un estudio de cohorte consiste en recoger al comienzo del estudio una sola muestra aleatoria de individuos de la población. Los individuos de esta muestra son encuestados para determinar quiénes están expuestos al factor de riesgo, y la muestra es entonces partida en función de esta información. Por ejemplo, si el interés es estudiar el efecto del consumo de cigarrillo sobre la incidencia de enfermedad coronaria del

*Profesor Asistente, Escuela de Estadística, Universidad Nacional de Colombia, Sede Medellín, Medellín, Colombia.

Correspondencia: Dr. Díaz, Escuela de Estadística, Universidad Nacional de Colombia, A.A. 3840, Medellín, Colombia. E-mail: fjdiaz@perseus.unalmed.edu.co

Artículo recibido: 26 de febrero de 2005; aceptado: abril 18 de 2005.

corazón, a los individuos de una muestra aleatoria puede preguntárseles si fuman o no fuman. Aquellos individuos sin enfermedad coronaria que fumen son asignados a un grupo, mientras que los individuos sin enfermedad coronaria que no fumen son asignados a un segundo grupo. Los dos grupos son entonces observados durante un período de tiempo, registrándose aquellos individuos que muestren la enfermedad durante ese período de tiempo. Luego se compara la aparición de la enfermedad coronaria en los dos grupos. La ventaja de este tipo de diseño es que también permite obtener la prevalencia del factor de riesgo en la población (por ejemplo, el porcentaje de fumadores en la población).

Nótese que en un estudio de cohorte pueden compararse más de dos grupos. Por ejemplo, en un estudio del fumar sobre enfermedades del corazón, los individuos pueden dividirse en fumadores actuales, ex fumadores e individuos que nunca han fumado.

Estudios de cohorte prospectivos versus estudios retrospectivos

Los diseños descritos hasta el momento reciben el nombre de prospectivos o longitudinales porque los investigadores están en contacto con los individuos desde el tiempo de referencia hasta el final del estudio, hasta que los individuos abandonan el estudio, o hasta que los individuos manifiestan la enfermedad. La principal ventaja de los estudios de cohorte prospectivos es que, cuando son bien controlados, se puede establecer una relación de causa-efecto entre el factor de riesgo y la enfermedad investigada.

Alternativamente, se pueden realizar estudios de cohorte retrospectivos, en donde la historia de los individuos se examina para reconstruir sus tiempos de supervivencia a cierto evento, y su exposición a cierto factor de riesgo. En los diseños retrospectivos, al igual que en los prospectivos, la expresión “tiempo de referencia” (*baseline*) se refiere al instante de tiempo a partir del cual se comienza a contar el tiempo de supervivencia. Los diseños retrospectivos tienen limitaciones debidas a la dificultad que los individuos pueden tener para recordar fechas y para recordar la presencia de factores de riesgo en su pasado. Además, puede ocurrir que algunos de los individuos que hayan estado sometidos a cierto factor de riesgo ya hayan muerto a causa de ese factor, lo cual impide que pertenezcan a la muestra que se está recogiendo actualmente, lo cual produce una subestimación del riesgo debido a ese factor.

Sin embargo, los estudios de cohorte retrospectivos son útiles en ciertos casos, por ejemplo, cuando el factor de riesgo sólo puede identificarse varios años después del tiempo de referencia. Es el caso de muchos factores de riesgo genéticos, los cuales se manifiestan fenotípicamente sólo años después de haber nacido el individuo (en este caso, la fecha de nacimiento sería el tiempo de referencia). Por

ejemplo, de León y colaboradores compararon las tasas de iniciación al consumo diario de cigarrillo de tres grupos de individuos:¹ pacientes esquizofrénicos, pacientes con desórdenes del humor y voluntarios de la población general de Lexington, Estados Unidos. Para ello, se le preguntó a los sujetos fumadores a qué edad empezaron a fumar diariamente. Luego, se hizo un análisis de supervivencia con dicha edad. En este caso, los factores de riesgo de interés fueron los dos tipos de enfermedad mental (esquizofrenia y desórdenes del humor), y el evento (enfermedad) de interés es el adquirir el hábito del cigarrillo. Como muchos de los pacientes comenzaron a fumar diariamente mucho antes de que se manifestaran sus primeros síntomas psiquiátricos, es claro que sólo un estudio de cohorte retrospectivo puede hacerse aquí. Las técnicas de análisis estadístico que describiremos en esta revisión, son aplicables tanto a estudios de cohorte prospectivos como retrospectivos.

Estudios de cohorte y análisis de supervivencia

Los estudios de cohorte también pueden clasificarse como estudios de cohorte fija o cohorte variable. En un estudio de cohorte fija, todos los sujetos entran al estudio al mismo tiempo y son seguidos durante la misma cantidad de tiempo. En este caso, la comparación de las morbilidades o mortalidades de los grupos puede hacerse usando técnicas estadísticas simples. Por ejemplo, se pueden calcular riesgos relativos o razones de odds. Por otro lado, en un estudio de cohorte variable, los sujetos entran al estudio en tiempos calendario distintos (tiempos de referencia móviles), o son sometidos a períodos de seguimiento distintos (períodos de seguimiento variables). En esta situación, se deben emplear técnicas de análisis de supervivencia. En este curso, nos limitaremos a estudios de cohorte variable. Véase Woodward para una excelente exposición de las técnicas usadas en estudios de cohorte fija y variable.²

En análisis de supervivencia, la palabra “sobrevivir” no necesariamente se refiere a no morir, sino que también se puede referir a no experimentar la enfermedad que se estudia. En general, sobrevivir significa no experimentar el evento de interés en el estudio. Un sujeto que no experimente el evento de interés antes de concluir el estudio se llama “censurado”, y los datos suministrados por dicho sujeto se llaman “datos censurados”. Así, aquellos sujetos que terminan el estudio en buen estado de salud son censurados.

Sujetos retirados y censura

Aquellos sujetos que abandonan el estudio antes de que este termine, y antes de que experimenten el evento de interés, se llaman “retiros” (*withdrawals*). El problema con estos sujetos es que nunca sabremos si experimentaron el evento de interés antes o después de terminarse el estudio.

Es importante notar que si los retiros son eliminados del análisis de una cohorte expuesta a un factor de riesgo, entonces el riesgo será sobreestimado porque no se está teniendo en cuenta que los sujetos que se retiraron fueron capaces de “sobrevivir” durante un período de tiempo antes de retirarse (es decir, no experimentaron la enfermedad antes de retirarse). Por otro lado, asumir que ninguno de los retiros hubiera experimentado la enfermedad hasta el final del estudio, lo cual es un supuesto muy común en análisis que usan razones de odds, implicaría subestimar el riesgo. Así, el análisis de supervivencia es la mejor alternativa cuando hay retiros. A no ser que el número de retiros sea pequeño con relación al tamaño muestral, las razones de odds sobreestimarán o subestimarán el riesgo.

El tratamiento más común de los retiros consiste en asumir que las causas de los retiros no están relacionadas con la enfermedad o evento en estudio, y entonces considerar a los retiros como censurados. Note que es posible que el retiro esté relacionado con la enfermedad, como cuando el sujeto abandona el estudio debido a que se está empezando a sentir mal, pero no se lo comunica a los investigadores (otro ejemplo, cuando el sujeto migra a un mejor clima, o cambia de trabajo o actividad para evitar el factor riesgo en estudio).

Una situación en la que un sujeto se trata como un retiro se presenta cuando el sujeto se enferma o muere por una causa distinta a la causa en estudio. Por ejemplo, si el interés es estudiar factores de riesgo de una enfermedad del corazón pero el sujeto muere o se retira por cáncer, ese sujeto se asume censurado. En este caso, se dice que el cáncer y la enfermedad del corazón son “riesgos en competencia”. Asumir como censurado un sujeto que se retira a causa de un riesgo en competencia es válido sólo cuando dicho riesgo es independiente del riesgo en estudio. En general, las técnicas más usuales del análisis de supervivencia asumen que cualquier tipo de censura ocurre independientemente del evento de interés (enfermedad o muerte).

Tablas de vida y funciones de supervivencia

Una tabla de vida es una presentación tabular del progreso de una cohorte a través del tiempo. El primer paso para construir una tabla de vida es dividir el período de seguimiento en intervalos consecutivos de tiempo. Luego se calculan las siguientes cantidades:

n_t = número de sobrevivientes en el tiempo t .

e_t = número de eventos que ocurren en el intervalo de tiempo que comienza en el tiempo t .

p_t = probabilidad de sobrevivir durante el intervalo de tiempo que comienza en el tiempo t .

q_t = probabilidad de que ocurra el evento de interés durante el intervalo que comienza en el tiempo t .

s_t = probabilidad de sobrevivir desde el comienzo del estudio hasta el final del intervalo que comienza en t (probabilidad de supervivencia acumulada).

Note que el tiempo de referencia (comienzo del estudio) es $t = 0$ y n_0 es entonces el tamaño muestral. Además, los intervalos de tiempo no tienen que tener la misma longitud. Observe que

$$\begin{aligned} q_t &= e_t / n_t, \\ p_t &= 1 - q_t, \\ s_t &= p_0 p_1 p_2 \dots p_t. \end{aligned}$$

Un problema de la fórmula anterior para s_t es que no tiene en cuenta el número de retiros en cada intervalo. Una manera de corregir este problema es calcular q_t con la fórmula alternativa

$$q_t = e_t / n_t^*$$

donde

$$n_t^* = n_t - c_t / 2, \text{ y}$$

c_t = número de sujetos censurados durante el intervalo que comienza en el tiempo t .

La lógica de la anterior fórmula se basa en que el número de sujetos que realmente estuvieron en riesgo durante el intervalo que comienza en el tiempo t es un número entre $n_t - c_t$ y n_t . Si todos los sujetos se retiraran al principio del intervalo, la probabilidad de que ocurra el evento de interés debería estimarse como $e_t / (n_t - c_t)$; pero si todos los sujetos se retiraran al final del intervalo, esta probabilidad debería estimarse como e_t / n_t . Por lo tanto, simplemente usamos un valor entre estos dos números. Este método se conoce como el método actuarial para datos de supervivencia, y asume que la censura ocurre uniformemente dentro de cada intervalo y que el riesgo de que ocurra el evento en un período de tiempo pequeño es constante a través de todo el intervalo (estos supuestos son razonables con intervalos pequeños).

El error estándar muestral para la probabilidad de supervivencia acumulada puede estimarse con la fórmula de Greenwood:

$$\hat{\sigma}(s_t) = s_t \sqrt{\sum_{i=0}^t \frac{q_i}{n_i^* - e_i}}$$

Usando una aproximación normal, un intervalo de confianza del 95% para una probabilidad de supervivencia acumulada particular es:

$$s_t \pm 1.96 \hat{\sigma}(s_t).$$

La figura 1 muestra una gráfica de s_t versus t para la edad de comienzo del fumar diariamente en los pacientes esquizofrénicos estudiados por de León y colaboradores.¹ Este tipo de gráfica se llama función (o curva) de supervivencia (*survival function*). Para calcular la función de supervivencia, de León y colaboradores implementaron el concepto de censura de la siguiente manera: Si un individuo de edad T nunca había fumado diariamente, su edad de comienzo de fumar diariamente fue considerada como censurada en T . Esto se hace para tener en cuenta el hecho de que el individuo puede comenzar a fumar en una edad posterior a T . Por otro lado, un individuo fumador no se consideró censurado si suministró la edad de comienzo de fumar diariamente. Finalmente, si un individuo fumaba (o había fumado) diariamente, no suministró la edad en que comenzó a fumar diariamente, pero suministró la edad en que fumó su primer cigarrillo, la edad de comienzo de fumar diariamente fue censurada en la edad del primer cigarrillo. Esta última estrategia es muy útil, especialmente porque algunas personas pueden tener dificultad para recordar la edad en que comenzaron a fumar diariamente.

En la función de supervivencia, a los puntos interiores de un intervalo se les asigna una probabilidad de supervivencia acumulada igual a la calculada para el extremo izquierdo del intervalo. Esto hace que la función sea escalonada. Como ejemplo, en la figura 1, para una t particular, la función de supervivencia me dice la probabilidad de llegar a la edad t sin adquirir el hábito del cigarrillo. La tabla de vida que proporcionó los cálculos para la figura 1 está dada en la tabla 1, la cual es una salida del paquete estadístico SPSS.³ Esta tabla de vida utiliza el método actuarial. Obsérvese que la probabilidad de supervivencia acumulada no puede crecer con el tiempo. En la tabla 2 se encuentran los errores estándar muestrales de las

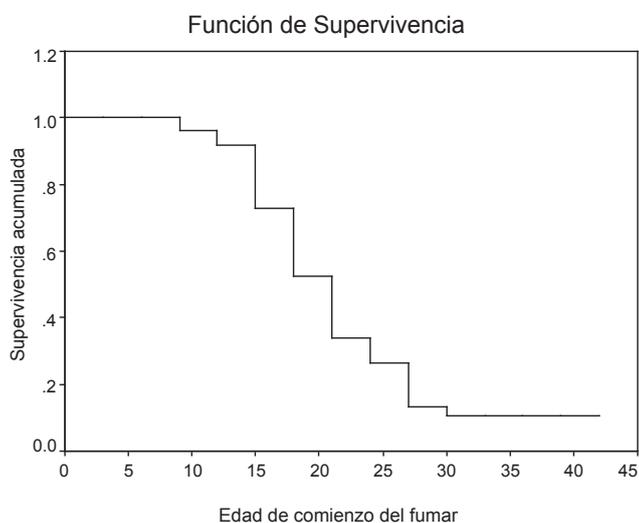


Figura 1. Probabilidades de supervivencia acumuladas versus la edad de comienzo de fumar diariamente en 50 pacientes esquizofrénicos. Las probabilidades fueron calculadas en una tabla de vida con intervalos de longitud igual a 3 años.

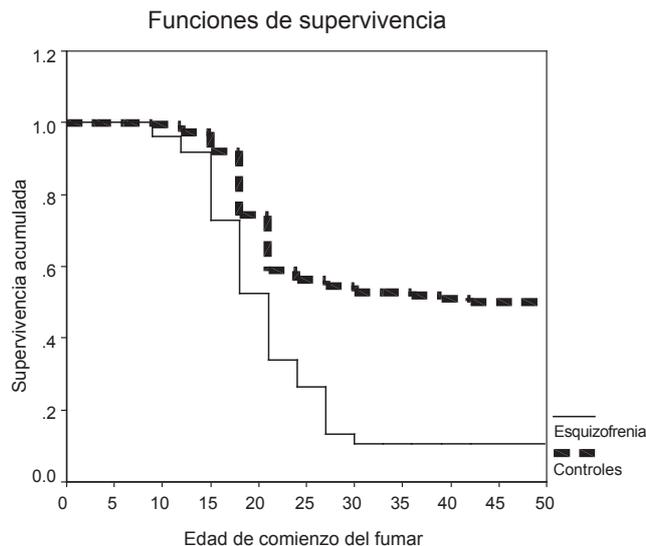


Figura 2. Probabilidades de supervivencia acumuladas versus la edad de comienzo de fumar diariamente en 50 pacientes esquizofrénicos y 410 voluntarios de la población general. Las probabilidades fueron calculadas en tablas de vida con intervalos de longitud igual a 3 años.

probabilidades de supervivencia acumuladas, calculados por SPSS con el método actuarial.

La figura 2 compara la función de supervivencia de los 50 pacientes esquizofrénicos con la función de supervivencia de 410 voluntarios de la población general de Lexington, E.U. En la figura 2, es interesante notar que las personas de la población general tienen una alta probabilidad de sobrevivir a la epidemia del cigarrillo, es decir, una alta proporción de personas que nunca adquieren el hábito de fumar diariamente. En general, a la proporción de sujetos que nunca contraerán la enfermedad de interés, se le llama proporción de inmunes. Una descripción de técnicas estadísticas para el análisis de la proporción de inmunes se encuentra en Maller y Zhou.^{4, 5}

Dos funciones de supervivencia pueden compararse en instantes de tiempo particulares, mediante el cálculo de intervalos de confianza o con pruebas de hipótesis. En efecto, supóngase que deseamos comparar dos probabilidades de supervivencia acumuladas en el tiempo t correspondientes a dos cohortes distintas. Por ejemplo, una cohorte estuvo expuesta al factor de riesgo mientras que la otra no lo estuvo. El error estándar muestral de la diferencia entre las dos probabilidades de supervivencia acumuladas, $s_t^{(1)}$ y $s_t^{(2)}$, puede estimarse con la fórmula

$$\hat{\sigma}(s_t^{(1)} - s_t^{(2)}) = \sqrt{\hat{\sigma}(s_t^{(1)})^2 + \hat{\sigma}(s_t^{(2)})^2}.$$

Así, un intervalo de confianza aproximado del 95% para la diferencia $s_t^{(1)} - s_t^{(2)}$ es

$$s_t^{(1)} - s_t^{(2)} \pm 1.96 \hat{\sigma}(s_t^{(1)} - s_t^{(2)}).$$

Tabla 1. Tabla de vida proporcionada por SPSS para la edad de comienzo del fumar diariamente en 50 pacientes esquizofrénicos. El evento de interés es comenzar a fumar diariamente.

t	n_t	c_t	n_t^*	e_t	q_t	p_t	s_t		
Intrvl Start Time	Number Entrng this Intrvl	Number Wdrawn During Intrvl	Number Exposed to Risk	Number of Termnl Events	Propn Terminating	Propn Surviving	Cumul Propn Surv at End	Proba-bility Densty	Hazard Rate
.0	50.0	.0	50.0	.0	.0000	1.0000	1.0000	.0000	.0000
3.0	50.0	.0	50.0	.0	.0000	1.0000	1.0000	.0000	.0000
6.0	50.0	1.0	49.5	2.0	.0404	.9596	.9596	.0135	.0137
9.0	47.0	1.0	46.5	2.0	.0430	.9570	.9183	.0138	.0147
12.0	44.0	2.0	43.0	9.0	.2093	.7907	.7261	.0641	.0779
15.0	33.0	1.0	32.5	9.0	.2769	.7231	.5250	.0670	.1071
18.0	23.0	1.0	22.5	8.0	.3556	.6444	.3384	.0622	.1441
21.0	14.0	1.0	13.5	3.0	.2222	.7778	.2632	.0251	.0833
24.0	10.0	.0	10.0	5.0	.5000	.5000	.1316	.0439	.2222
27.0	5.0	.0	5.0	1.0	.2000	.8000	.1053	.0088	.0741
30.0	4.0	.0	4.0	.0	.0000	1.0000	.1053	.0000	.0000
33.0	4.0	.0	4.0	.0	.0000	1.0000	.1053	.0000	.0000
36.0	4.0	.0	4.0	.0	.0000	1.0000	.1053	.0000	.0000
39.0	4.0	.0	4.0	.0	.0000	1.0000	.1053	.0000	.0000
42.0+	4.0	4.0	2.0	.0	.0000	1.0000	.1053	**	**

** These calculations for the last interval are meaningless.
The median survival time for these data is 18.40

En caso de que se quiera examinar la hipótesis nula de que $s_t^{(1)}$ y $s_t^{(2)}$ son iguales, se puede usar el estadístico de prueba

$$\left\{ \frac{s_t^{(1)} - s_t^{(2)}}{\hat{\sigma}(s_t^{(1)} - s_t^{(2)})} \right\}^2,$$

el cual tiene aproximadamente una distribución chi-cuadrada con 1 grado de libertad, cuando los tamaños muestrales son suficientemente grandes.

Una limitación de un procedimiento de prueba de hipótesis que use el anterior estadístico de prueba, es que no permite examinar la hipótesis nula global de que dos curvas de supervivencia son iguales en *todos* los instantes de tiempo. Tal procedimiento sólo permitiría comparar dos curvas de supervivencia en instantes de tiempo *particulares*. Además, comparar dos curvas de supervivencia mediante la aplicación repetida de dicho procedimiento en muchos instantes de tiempo, tiene como consecuencia indeseable que se tendría una alta probabilidad de rechazar la hipótesis nula, incluso siendo falsa. Así, tarde que temprano terminaríamos concluyendo que las dos curvas son estadísticamente distintas, así no sea cierto. Una solución a este problema es usar pruebas globales para la comparación de dos curvas de supervivencia, las cuales se presentarán más adelante.

Tabla 2. Resultados de SPSS que muestran los errores estándar muestrales de las probabilidades de supervivencia acumuladas de la Tabla 1, para los datos de edad de comienzo del fumar diariamente en pacientes esquizofrénicos.

t	$\hat{\sigma}(s_t)$		
Intrvl Start Time	SE of Cumul Surviving	SE of Proba-bility Densty	SE of Hazard Rate
.0	.0000	.0000	.0000
3.0	.0000	.0000	.0000
6.0	.0280	.0093	.0097
9.0	.0391	.0095	.0104
12.0	.0648	.0192	.0258
15.0	.0738	.0199	.0353
18.0	.0712	.0197	.0498
21.0	.0673	.0138	.0477
24.0	.0535	.0178	.0937
27.0	.0489	.0086	.0736
30.0	.0489	.0000	.0000
33.0	.0489	.0000	.0000
36.0	.0489	.0000	.0000
39.0	.0489	.0000	.0000
42.0+	.0489	**	**

El estimador de Kaplan-Meier

Uno de los estimadores de la función de supervivencia más utilizados es el estimador de Kaplan-Meier (KM). En esencia, este estimador se halla construyendo una tabla de vida con los intervalos más pequeños posibles. En efecto, suponga que el evento de interés ocurre en los D tiempos distintos $t_1 < t_2 < \dots < t_{D-1} < t_D$. El estimador KM es una curva de supervivencia calculada en la forma descrita en la sección 2, usando los intervalos semiabiertos $[t_1, t_2), [t_2, t_3), \dots, [t_{D-1}, t_D)$. Sin embargo, el estimador de KM no realiza ninguna corrección debida a censura. Es decir, el estimador de KM usa simplemente n_i en vez de n_i^* , tanto para el cálculo de q_i como para el cálculo de $\hat{G}(s_i)$. Esto se debe a que con intervalos pequeños la corrección por censura no produce cambios notorios en los cálculos. Además, cuando el número de sujetos en el estudio es relativamente grande y las censuras ocurren independientemente del evento de interés, el estimador de KM se aproxima razonablemente a la función de supervivencia poblacional. Una ventaja de emplear los intervalos más pequeños posibles es que se utiliza la máxima cantidad de información proporcionada por los datos. Sin embargo, el despliegue tabular obtenido ya no es tan útil.

Debe notarse que el supuesto de intervalos pequeños, necesario para el cálculo del estimador de KM, podría no ser válido si los sujetos investigados son monitoreados a intervalos muy grandes de tiempo.

La tabla 3 muestra resultados de un análisis de Kaplan-Meier de SPSS para la edad de comienzo del fumar diariamente en los pacientes esquizofrénicos. Obsérvese que el despliegue tabular es muy grande, lo cual es el resultado de trabajar con los intervalos más pequeños posibles. La figura 3 muestra el estimador de Kaplan-Meier de la función de supervivencia. Los círculos negros sobre la curva muestran los puntos donde se presentaron censuras.

Función de riesgo acumulado

En muchas situaciones, dibujar la función de riesgo acumulado (*cumulative hazard function*) es más útil e informativo que dibujar la función de supervivencia. Usando la notación de las secciones 2 y 3, la función de riesgo acumulado se estima con la fórmula de Nelson-Aalen

$$H(t) = \sum_{t_i \leq t} q_i .$$

Al igual que con el estimador de Kaplan-Meier, no se hacen correcciones por censura, aunque esto afecta poco las conclusiones cuando se trabaja con grupos relativamente grandes de individuos e intervalos de monitoreo pequeños.

En la práctica, lo que se observa en una gráfica de la función de riesgo acumulado es la forma como cambia su

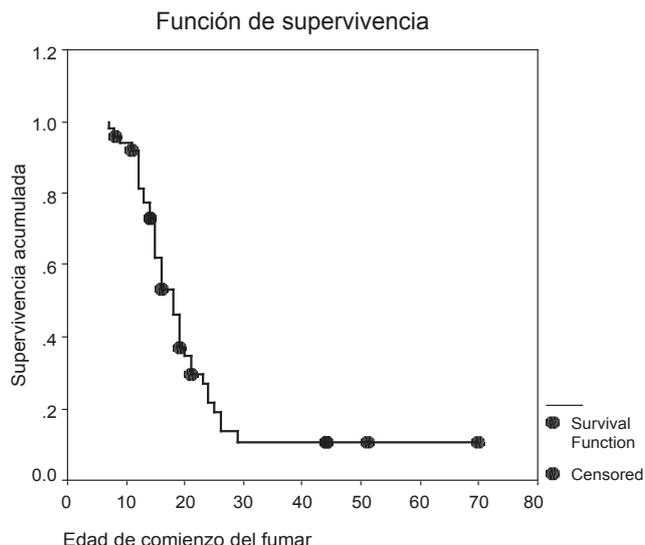


Figura 3. Probabilidades de supervivencia acumuladas versus la edad de comienzo de fumar diariamente en 50 pacientes esquizofrénicos. Las probabilidades fueron calculadas usando el método de Kaplan-Meier.

pendiente a través del tiempo. Esto es un indicativo de la forma como cambia el riesgo de que ocurra la enfermedad para un individuo particular a través del tiempo. En efecto, para un instante de tiempo t , la pendiente de la función de riesgo acumulado se interpreta como la tasa de riesgo (*hazard rate*) para ese instante de tiempo. La tasa de riesgo para un instante de tiempo particular es, por definición, la proporción por unidad de tiempo de individuos que nunca sufrieron de la enfermedad antes de ese instante de tiempo y que comenzarán a sufrir de la enfermedad en ese instante.

La figura 4 muestra la función de riesgo acumulado para la edad de comienzo del fumar diariamente en los 410 voluntarios de la población general. Una ventaja de usar este tipo de gráfica es que puede examinarse el cambio en las tasas de iniciación al fumar diariamente a través de las distintas edades.^{1,5} La tasa de iniciación al fumar para una edad particular se define como la tasa de riesgo para esa edad. A su vez, por definición, la tasa de riesgo para una edad particular es la proporción por unidad de tiempo de las personas que nunca fumaron diariamente antes de esa edad particular y que comenzarán a fumar diariamente en esa edad específica. La tasa de iniciación al fumar para una edad particular es la pendiente de la curva (función) de riesgo acumulado en esa edad. Entre más pendiente sea la curva de riesgo acumulado en una edad particular, más alta es la tasa de iniciación para esa edad.

Esta figura 4 muestra que las tasas de iniciación al fumar más altas se encontraron en el grupo de los individuos que están entre los 15 y los 20 años de edad. Entre los 20 y los

Tabla 3. Resultados de SPSS que muestran el cálculo de las probabilidades de supervivencia acumuladas para la edad de comienzo del fumar diariamente en 50 pacientes esquizofrénicos, usando el método de Kaplan-Meier.

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
7	uncensored	.9800	.0198	1	49
8	uncensored	.9600	.0277	2	48
8	right censored			2	47
9	uncensored	.9396	.0338	3	46
11	uncensored	.9191	.0388	4	45
11	right censored			4	44
12	uncensored			5	43
12	uncensored			6	42
12	uncensored			7	41
12	uncensored			8	40
12	uncensored	.8147	.0558	9	39
13	uncensored			10	38
13	uncensored	.7729	.0603	11	37
14	uncensored			12	36
14	uncensored	.7311	.0638	13	35
14	right censored			13	34
14	right censored			13	33
15	uncensored			14	32
15	uncensored			15	31
15	uncensored			16	30
15	uncensored			17	29
15	uncensored	.6204	.0708	18	28
16	uncensored			19	27
16	uncensored			20	26
16	uncensored			21	25
16	uncensored	.5317	.0733	22	24
16	right censored			22	23
18	uncensored			23	22
18	uncensored			24	21
18	uncensored	.4624	.0738	25	20
19	uncensored			26	19
19	uncensored			27	18
19	uncensored			28	17
19	uncensored	.3699	.0721	29	16
19	right censored			29	15
20	uncensored	.3452	.0714	30	14
21	uncensored			31	13
21	uncensored	.2959	.0692	32	12
21	right censored			32	11
23	uncensored	.2690	.0679	33	10
24	uncensored			34	9
24	uncensored	.2152	.0641	35	8
25	uncensored	.1883	.0615	36	7
26	uncensored			37	6
26	uncensored	.1345	.0544	38	5
29	uncensored	.1076	.0498	39	4
44	right censored			39	3
44	right censored			39	2
51	right censored			39	1
70	right censored			39	0

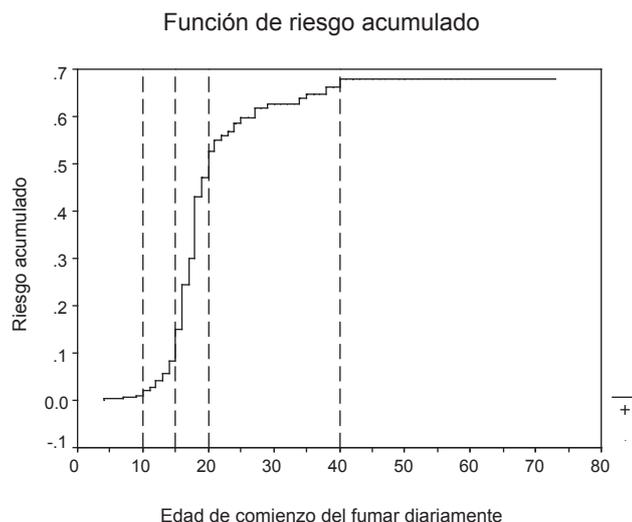


Figura 4. Función de riesgo acumulado* para la edad de comienzo del fumar diariamente en 410 voluntarios de la población general. (*Cumulative hazard function).

40 años de edad, las tasas de iniciación al fumar disminuyeron bruscamente. Finalmente, las tasas de iniciación al fumar entre los individuos mayores de 40 años de edad fueron prácticamente nulas. Además, a partir de los 10 años de edad, las tasas de iniciación al fumar crecieron de manera acelerada hasta los 15 años, edad en que se alcanza una tasa máxima que se mantuvo estable hasta los 20 años.

A nivel de un solo individuo, la figura 4 se puede interpretar de la siguiente forma: El riesgo de que un individuo de la población general empiece a fumar diariamente comienza a aumentar de manera acelerada a partir de los 10 años, hasta los 15 años. A los 15 años el individuo alcanza el máximo riesgo, el cual se mantiene constante hasta los 20 años. Después de los 20 años el riesgo decrece de manera rápida hasta los 40 años. Finalmente, es muy improbable que un individuo mayor de 40 años de edad, que nunca ha fumado diariamente, adquiera el hábito de fumar diariamente.

Además de la utilidad de la función de riesgo acumulado para explorar la evolución del riesgo a través del tiempo, esta función es útil para identificar un modelo paramétrico adecuado para los datos, ya que caracteriza de manera unívoca la función de supervivencia.

Comparación de dos funciones de supervivencia

A continuación se presentan procedimientos de prueba para examinar la hipótesis nula global de que dos curvas de supervivencia son iguales en *todos* los instantes de tiempo que fueron abarcados en el período de seguimien-

to. Debe notarse que tal hipótesis nula es equivalente a la hipótesis nula de que las dos funciones de riesgo acumulado son iguales. Esto se debe a que las curvas de supervivencia determinan de manera única las funciones de riesgo acumulado y viceversa.

Prueba de Cochran-Mantel-Haenszel. Supóngase que se desea comparar las dos curvas de supervivencia de dos cohortes. Una cohorte está expuesta al factor de riesgo, mientras que la otra no lo está (cohorte de control). Al igual que cuando se construyen dos tablas de vida para propósitos de comparación, el primer paso para realizar el procedimiento de prueba de Cochran-Mantel-Haenszel es dividir de manera similar los períodos de seguimiento de las dos cohortes en intervalos consecutivos de tiempo. Luego, para el intervalo que comienza en el tiempo t , se construye la tabla de contingencia 2×2 mostrada en la tabla 4. Realizando una tabla como esta para todos los intervalos, se procede a calcular el estadístico de prueba de Cochran-Mantel-Haenszel

$$\frac{\left(\sum_t (e_t^{(1)} - E(e_t^{(1)})) \right)^2}{\sum_t V(e_t^{(1)})}$$

donde

$$E(e_t^{(1)}) = \frac{m_t n_t^{(1)}}{n_t^{(1)} + n_t^{(2)}}$$

$$V(e_t^{(1)}) = \frac{m_t \bar{m}_t n_t^{(1)} n_t^{(2)}}{\left(n_t^{(1)} + n_t^{(2)} \right)^2 \left(n_t^{(1)} + n_t^{(2)} - 1 \right)}$$

En esencia, el estadístico de Cochran-Mantel-Haenszel prueba de manera simultánea la hipótesis de independencia para cada una de las tablas de contingencia construidas. De esta forma, examina la hipótesis nula de que el factor de riesgo y la ocurrencia de la enfermedad son independientes. Bajo esta hipótesis, con muestras suficientemente grandes, este estadístico se distribuye aproximadamente como una chi-cuadrado con 1 grado de libertad.

La prueba log-rank. El procedimiento de prueba de Cochran-Mantel-Haenszel tiene la desventaja de que no hace ninguna corrección por censura. Si los intervalos construidos son anchos, esto puede originar un sesgo en las conclusiones. Por esta razón, se prefieren utilizar los intervalos de tiempo más pequeños posibles, al igual que con el estimador de KM. En este caso, sin embargo, se combinan los intervalos más pequeños proporcionados por las dos cohortes. El procedimiento de prueba así obtenido se llama prueba log-rank, y, en este caso, el estadístico de prueba se llama estadístico log-rank.

Debe notarse que la prueba log-rank asume que los riesgos de sufrir la enfermedad por parte de las dos cohortes son proporcionales (*proportional hazards*). Es decir, asume que el cociente de los dos riesgos es una constante que no depende del tiempo. Cuando hay evidencia de que este supuesto es falso, se prefiere utilizar una prueba log-rank ponderada, la cual usa el estadístico de prueba

$$\frac{\left(\sum_t w_t (e_t^{(1)} - E(e_t^{(1)})) \right)^2}{\sum_t w_t^2 V(e_t^{(1)})}$$

donde $\{w_t\}$ es un conjunto de ponderaciones (*weights*), uno para cada intervalo. Bajo la hipótesis nula de que el factor de riesgo y la ocurrencia de la enfermedad son independientes, y con muestras suficientemente grandes, este estadístico sigue teniendo aproximadamente una distribución chi-cuadrado con 1 grado de libertad.

Cuando $w_t = \frac{n_t^{(1)} + n_t^{(2)}}{n_t^{(1)}}$, es decir, cuando w_t es el número total de individuos en riesgo en el instante de tiempo t , la prueba log-rank se conoce como la prueba de Breslow, o la prueba de Wilcoxon generalizada. Esta prueba le da mayor peso a los intervalos de tiempo con mayor número de individuos en riesgo, los cuales son inevitablemente los primeros intervalos. Otra prueba ponderada es la de Tarone-Ware, la cual utiliza las ponderaciones

$$w_t = \sqrt{\frac{n_t^{(1)} + n_t^{(2)}}{n_t^{(1)}}}$$

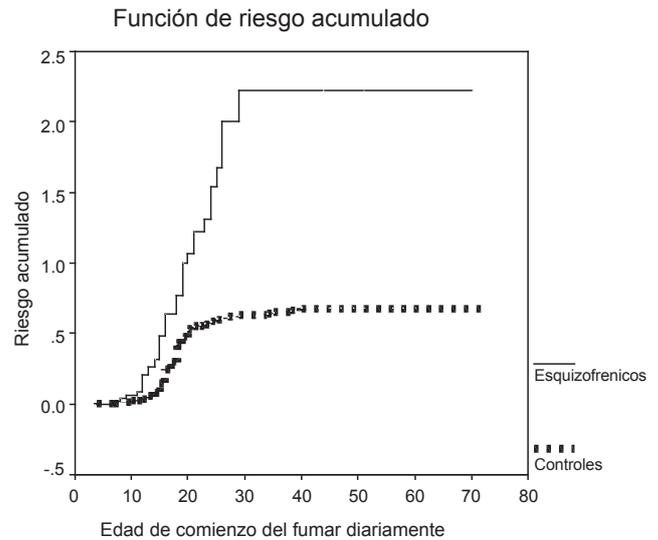


Figura 5. Función de riesgo acumulado para la edad de comienzo del fumar diariamente en pacientes esquizofrénicos y voluntarios de la población general (controles).

Como ilustración, obsérvese la figura 5, en donde se comparan las funciones de riesgo acumulado para pacientes esquizofrénicos y voluntarios de la población general.¹ para una descripción de la censura en las muestras, véase la tabla 5. La figura 5 nos permite comparar las tasas de iniciación al fumar diariamente de las dos cohortes. Esta figura sugiere que las dos cohortes tuvieron esencialmente

Tabla 4. Tabla de contingencia 2x2 correspondiente al intervalo de tiempo que comienza en t , usada para calcular el estadístico de Cochran-Mantel-Haenszel. Este estadístico se usa para comparar las funciones de supervivencia de dos cohortes, una de las cuales está expuesta a un factor de riesgo pero la otra no.

Exposición al Factor de riesgo	Ocurrencia del evento o enfermedad		
	Ocurrió	No ocurrió	Total
Expuesta	$e_t^{(1)}$	$n_t^{(1)} - e_t^{(1)}$	$n_t^{(1)}$
No expuesta	$e_t^{(2)}$	$n_t^{(2)} - e_t^{(2)}$	$n_t^{(2)}$
Total	m_t	\bar{m}_t	$n_t^{(1)} + n_t^{(2)}$

$e_t^{(1)}$ = Número de eventos que ocurren en el intervalo que comienza en el tiempo t , correspondiente a los sujetos que están expuestos al factor de riesgo.

$n_t^{(1)}$ = Número de sobrevivientes en el tiempo t , correspondiente a los sujetos que están expuestos al factor de riesgo.

$e_t^{(2)}$ = Número de eventos que ocurren en el intervalo que comienza en el tiempo t , correspondiente a los sujetos que no están expuestos al factor de riesgo.

$n_t^{(2)}$ = Número de sobrevivientes en el tiempo t , correspondiente a los sujetos que no están expuestos al factor de riesgo.

Tabla 5. Descripción de la censura en la edad de comienzo del fumar diariamente en tres muestras de sujetos: Pacientes esquizofrénicos, pacientes con desórdenes del humor y voluntarios de la población general (controles).

	Total eventos	Número censurados	Número censurados	Porcentaje
Controles	410	189	221	53.90
Esquizofrénicos	50	39	11	22.00
Desórdenes del humor	41	25	16	39.02
Muestras juntas	501	253	248	49.50

las mismas tasas de iniciación antes de los 20 años de edad. Sin embargo, de los 20 a los 29 años, los pacientes con esquizofrenia tuvieron tasas de iniciación más altas que los individuos de la población general. Más aún, cuando los pacientes esquizofrénicos tuvieron entre 20 y 29 años, sus tasas de iniciación al fumar fueron iguales que cuando tuvieron entre 15 y 20 años. Finalmente, obsérvese que a partir de los 29 años, los esquizofrénicos y las personas de la población general tuvieron tasas de iniciación muy bajas.¹ Esto último sugiere que después de los 29 años fue muy improbable adquirir el hábito de fumar diariamente.

Se concluye que, aproximadamente a partir de los 20 años, las personas con vulnerabilidad a la esquizofrenia tienden a tener una probabilidad más alta de adquirir el hábito del cigarrillo que las personas de la población general. Sin embargo, esto sólo ocurre hasta los 29 años aproximadamente.

Para verificar la observación de que las tasas de iniciación al fumar diariamente de las dos cohortes son diferentes dentro de algunos grupos de edad, de León y colaboradores realizaron pruebas log-rank.¹ La tabla 6 muestra que las tasas de iniciación son significativamente distintas en al menos un grupo de edad, a un nivel de significación del 5%. O, dicho de otra forma, las dos funciones de riesgo acumulado son significativamente distintas.

Prueba log-rank estratificada. La prueba log-rank permite hacer ajustes para evitar los efectos indeseados de variables de confusión. Una variable de confusión es

Tabla 6. Resultados de SPSS mostrando las pruebas para contrastar la hipótesis nula de igualdad de las funciones de riesgo acumulado de los pacientes esquizofrénicos y los controles (ver figura 5).

Test Statistics for Equality of Survival Distributions for age started daily smoking			
	Statistic	df	Significance
Log Rank	34.10	1	.0000
Breslow	27.86	1	.0000
Tarone-Ware	30.56	1	.0000

una variable que puede explicar completa o parcialmente la relación entre el factor de riesgo y la enfermedad, y que por lo tanto le hace creer a uno que el factor de riesgo es lo que explica las diferencias observadas entre las curvas de supervivencia, sin ser así. Una variable de confusión también puede enmascarar el efecto de un factor de riesgo. Para hacer ajustes por variables de confusión, los sujetos se dividen de acuerdo a los estratos definidos por los niveles de las variables de confusión. Por ejemplo, si se tienen dos variables de confusión, cada una con tres niveles, entonces los sujetos se dividen de acuerdo a cada una de las 3×3=9 combinaciones de niveles de las dos variables. Cada una de estas combinaciones sería un estrato. A continuación, el numerador y el denominador del estadístico log-rank se calculan para cada estrato. Sean E_j y V_j el numerador y denominador del estadístico log-rank, respectivamente, calculados sólo con los datos de los sujetos del estrato j -ésimo. Entonces, el estadístico log-rank, ajustado por variables de confusión, es

$$\sum_j E_j / \sum_j V_j ,$$

donde las sumas se hacen a través de todos los estratos. Este estadístico también tiene una distribución aproximada chi-cuadrado con un grado de libertad, cuando la hipótesis nula de independencia del factor de riesgo y la enfermedad es correcta, y cuando el tamaño muestral es suficientemente grande.

Como ilustración, los pacientes esquizofrénicos y los controles de la figura 5 fueron comparados usando una prueba log-rank estratificada que ajustó por los efectos de las variables sexo y nivel de educación.¹ La razón por la cual se hizo este ajuste es que en análisis previos se había observado que existían diferencias significativas entre las tasas de iniciación al fumar de hombres y mujeres, y entre las tasas de iniciación de personas con nivel educativo alto y bajo. Los resultados están en la tabla 7, los cuales sugieren que las diferencias observadas entre las funciones de riesgo acumulado de los pacientes y de los controles

Tabla 7. Resultados de SPSS mostrando las pruebas para contrastar la hipótesis nula de igualdad de las funciones de riesgo acumulado de los pacientes esquizofrénicos y los controles, estratificando por sexo y nivel de educación.

Test Statistics for Equality of Survival Distributions for age started daily smoking Adjusted for gender and education			
	Statistic	df	Significance
Log Rank	12.18	1	.0005
Breslow	6.76	1	.0093
Tarone-Ware	9.15	1	.0025

no son explicadas por distribuciones desiguales del sexo o del nivel de educación en las dos poblaciones.

El método persona-años

Un método alternativo para el análisis de datos censurados es el método persona-años. En esencia, para cada cohorte en estudio, este método calcula la tasa de incidencia persona-años (*person-years incidence rate*). Esta tasa está definida como

$$\hat{\rho} = e / y,$$

donde e es el número total de veces que se presenta la enfermedad (o evento de interés) en la cohorte, e y es la suma de todos los períodos de seguimiento de los sujetos de la cohorte. El período de seguimiento de un sujeto termina cuando el sujeto es censurado o cuando manifiesta la enfermedad. La cantidad y se llama el número de persona-años de observación (o de seguimiento). Así, $\hat{\rho}$ se interpreta como el número de eventos por año de observación.

Con muestras grandes, un intervalo de confianza del 95% para $\hat{\rho}$ está dado por la fórmula (ρ_L, ρ_U), en donde

$$\rho_L = \left(\frac{1.96}{2} - \sqrt{e} \right)^2 / y,$$

$$\rho_U = \left(\frac{1.96}{2} + \sqrt{e+1} \right)^2 / y.$$

Debe notarse que este intervalo de confianza asume que los eventos de enfermedad ocurren aleatoria e independientemente, y que e tiene una distribución de Poisson. En muchas situaciones reales, estos supuestos se pueden considerar razonables.

Como ilustración, la tasa de incidencia del consumo diario de cigarrillo para los pacientes esquizofrénicos, multiplicada por 1000, fue 40 por 1.000 persona-años (intervalo de confianza = (29, 55)), mientras que la tasa de incidencia para los controles fue 18 por 1.000 persona-años (intervalo de confianza = [16, 21]). Estos intervalos de confianza pueden calcularse con el paquete de computador CIA.⁶ Estas dos tasas pueden compararse calculando la tasa relativa, 40/18=2.2. Un intervalo de confianza del 95% para la tasa relativa es (1.5, 3.0).² Este último intervalo de confianza sugiere que la tasa de incidencia de los pacientes esquizofrénicos fue significativamente mayor que la de los controles.

A menudo, las tasas de incidencia persona-años son corregidas por la edad, especialmente cuando los sujetos comienzan a ser monitoreados (prospectiva o retrospectivamente) a edades diferentes. Véase Woodward para mayores detalles.²

Summary

The cohort design is a major methodological tool in epidemiological and medical research. This article introduces the basic concepts and statistical methods supporting this design. After describing the fundamental ideas of prospective and retrospective cohort designs and of fixed and variable cohort designs, this article focuses on the basic tools used in variable cohort designs. These tools are part of the so called survival analysis. The concepts of censoring, life table, survival function, the Kaplan-Meier estimator and the cumulative hazard function are then described. Statistical methods for comparing two survival functions such as the Cochran-Mantel-Haenszel and log-rank tests and the log-rank stratified test are also explained. Finally, the alternative person-years method is introduced. The concepts and methods are illustrated with a retrospective cohort study of the age of onset of daily smoking in patients with schizophrenia.

Key words: Cohort design, survival analysis, survival curves, Kaplan-Meier estimator, cumulative hazard function, Cochran-Mantel-Haenszel test, log-rank test, person-years method, smoking, schizophrenia.

Agradecimientos

El autor agradece al Dr. José de León, director del Centro de Investigaciones en Salud Mental de la Universidad de Kentucky, E.U., por proporcionarle los datos de edad de comienzo del fumar.

Este artículo fue presentado como un curso corto en el XIII Simposio de Estadística, organizado por la Universidad Nacional de Colombia y la Universidad del Quindío y realizado del 3 al 7 de agosto del 2003 en Armenia, Quindío, Colombia.

Referencias

1. De León J, Diaz FJ, Rogers T, Browne D, Dinsmore L. Initiation of daily smoking and nicotine dependence in schizophrenia and mood disorders. *Schizophrenia Research* 2002; 56:47-54.
2. Woodward M. *Epidemiology, study design and data analysis*. Boca Raton, Chapman & Hall/CRC, 1999.
3. SPSS Advanced Models 12.0. Chicago, SPSS Inc, 2003.
4. Maller RA, Zhou S. Estimating the proportion of immunes in a censored sample. *Biometrika* 1992; 79:731-9.
5. Maller RA, Zhou S. Testing for the presence of immune or cured individuals in censored survival data. *Biometrics* 1995; 51:1197-205.
6. Gurpegui M, Martínez-Ortega JM, Aguilar MC, Díaz FJ, Quintana HM, de León J. Smoking initiation and schizophrenia: A replication study in a Spanish sample. *Schizophrenia Research* 2005 (in press).
7. Altman DG, Machin D, Bryant TN, Gardner, MJ. *Statistics with confidence. Confidence intervals and statistical guidelines*. Bristol, BMJ Books, 2000.