APLICACIÓN DEL MODELO DE RIESGOS PROPORCIONALES DE COX A PACIENTES CON SIDA EN ESPAÑA

E. Corpas-Nogales*[1] and A. M. Lara-Porras[2]**
, Andalusian School of Public Health, Granada.
, Department of Statístics and Operations Research. Faculty of Sciences. University of Granada. Spain

ABSTRACT
In the last years there has been a rapid development of probability models and statistical analysis for technological and medical survival data. The proportional hazard model formulated by Cox (1972) has rapidly become one of the most widely used methods for the analysis of clinical data, turning into a very powerful tool that has served to generalize the use of the analysis of survival This model is particulary useful examining treatment comparisons based on the times to some event while adjusting for the effects of concomitant variables and easily adapts to the incomplete data which are frequent in medical research. We work with censored data on the right, for which we used the statistical package S-Plus.
In this paper we applied the Cox´s proportional hazard model in a study of 1366 women infected with AIDS in Andalusia (Spain) between the years 1982 and 2001 focusing on the following variables: age, time of diagnosis, province of residence and level of transmission.
The main result has been an increase in the survival of women with the introduction of the "Antiretroviral Therapy Very Active" in 1996, women diagnosed in the period 1993-1995 have an irrigation die 2.3 times greater than that diagnosed after 1996. As for age, women 35 years and older have a risk of death 1.8 times higher than in women under 30. The province with the largest survival is Grenada. We have not found any differences in the category of transmission. The data has been provided by the Andalusian Registry of AIDS.

KEY WORDS: Cox´s model, survival, AIDS.

MSC: 62N01; 62N02; 62P10

RESUMEN
En los últimos años se ha producido un rápido desarrollo de los modelos probabilísitcos y de análisis estadístico aplicados a datos de supervivencia. El modelo de riesgos proporcionales formulado por Cox se ha convertido en uno de los métodos más utilizados, siendo los temas relacionados con la salud una de las áreas en las que se ha obtenido mayor aplicación. Se ha convertido en una herramienta muy potente que ha servido para generalizar el uso del análisis de supervivencia. Este modelo es particularmente útil para comparar grupos en los que se estudia el tiempo transcurrido hasta la ocurrencia de un evento, pudiendo analizar conjuntamente el efecto de varias variables. El modelo de Cox se adapta fácilmente a situaciones con datos incompletos, algo que ocurre muy frecuentemente en investigación. En nuestro estudio trabajamos con datos censurados a la derecha. Los datos se han analizado con el paquete estadístico S-Plus.
Se ha aplicado el modelo de Cox a datos de mujeres infectadas de SIDA. Se han estudiado 1366 mujeres diagnosticadas en Andalucía entre 1982 y 2001. Las variables utilizadas para construir el modelo han sido: edad, fecha de diagnóstico, provincia de residencia y categoría de transmisión. Los datos han sido proporcionados por el Registro Andaluz de Casos de Sida.
En este estudio se ha observado que la introducción del TARGA mejora la supervivencia de las mujeres diagnosticadas de SIDA, las mujeres diagnosticadas en el periodo 1993-1995 tienen un riego de morir 2.3 veces superior que las diagnosticadas después de 1996. En cuanto a la edad, las mayores de 35 años tienen un riesgo de morir 1.8 veces superior que las menores de 30. La provincia con mayor supervivencia es Granada. No se han encontrado diferencias en cuanto a la categoría de transmisión.

1. INTRODUCCIÓN

El SIDA se ha convertido en la última epidemia del siglo XX (Pachón, 2000). El SIDA es la etapa final y más grave de la enfermedad del Virus de la Inmunodeficiencia Humana, una infección viral causada por el virus

[1] ecorpas@yahoo.es
[2] alara@ugr.es

de la inmunodeficiencia humana que destruye de forma gradual el sistema inmune, causando infecciones difíciles de combatir por el cuerpo humano.

El primer caso de SIDA fue diagnosticado en Estados Unidos el 6 de junio de 1981. En España, el primer caso, se diagnosticó en Barcelona también en el año 1981. A pesar de que los avances terapéuticos han modificado el desarrollo de esta enfermedad en los países desarrollados, el SIDA se encuentra lejos de estar bajo control, convirtiéndola en una de las enfermedades más graves de nuestros días.

Con este trabajo nos planteamos estudiar el tiempo de supervivencia de las mujeres diagnosticadas de SIDA en Andalucía, analizando cuales son los factores asociados a un mayor tiempo de supervivencia entre las pacientes con esta enfermedad. Para este propósito se han utilizado los datos del Registro Andaluz de Casos de Sida y hemos considerado el modelo de riesgos proporcionales introducido por Cox (1972).

Este artículo consta de una sección en la que se presentan diferentes modelos de supervivencia, el estimador de Kaplan-Meier y el modelo de Cox, así como la propiedad de riesgos proporcionales necesaria para poder utilizar el modelo de Cox. En otra sección se presentan los resultados del análisis de supervivencia en las mujeres infectadas de SIDA en Andalucía. Se estudia el tiempo supervivencia de estos sujetos y los factores que pueden estar asociados a dicha supervivencia.

2. ANÁLISIS DE SUPERVIVENCIA

2.1. Modelos de supervivencia

Los modelos de supervivencia han sido muy utilizado a lo largo de la historia, siendo los temas relacionados con la salud una de las áreas en las que se ha obtenido mayor aplicación, especialmente en áreas relativas a enfermedades crónicas y enfermedades de alta letalidad.

Un método muy utilizado en análisis de supervivencia es el método de Kaplan- Meier (Collet, 1994), que obtiene la supervivencia de forma recursiva mediante la siguiente expresión

$$S(t_i) = \frac{r_i - m_i}{r_i} S(t_{i-1})$$

donde r_i son los individuos a riesgo en el instante i y m_i son los individuos que han presentado el evento en el instante de tiempo i. Sin embargo, este método no permite plantear un modelo de regresión para la supervivencia en función de variables explicativas.

Con el método de Kaplan-Meier se asume que los grupos son homogéneos con respecto a todas las variables, pero esto no es siempre así. Existen diferentes alternativas para construir modelos de supervivencia en los que se tenga en cuenta la relación de diferentes variables medidas en cada individuo, sin embargo, el método más utilizado es el de riesgos proporcionales de Cox (Cox, 1984). La función de riesgo para el modelo de Cox viene dada por la ecuación

$$\lambda_i(t) = \lambda_0(t) e^{X_i(t)\beta}$$

donde $\lambda_0(t)$ es una función 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. El método de Cox se utiliza fundamentalmente para comparar grupos, calculando cocientes de riesgo.

La estimación de β se basa en la función de verosimilitud parcial introducida por Cox (1972). El estimador obtenido, $\hat{\beta}$, es consistente y tiene distribución asintótica normal con media el verdadero vector de parámetros, β, y varianza la inversa de la matriz de información esperada, $\left\{E\left[I(\beta)\right]\right\}^{-1}$

2.2. Censuras

Consideremos T_i el tiempo desde el comienzo del estudio hasta que ocurre el evento de interés para el i-ésimo paciente. En ocasiones el tiempo de seguimiento es limitado, y no es lo suficientemente largo como para que en todos los sujetos del estudio se observe el evento de interés. En este caso, a los sujetos en los que no se ha podido observar el evento se les llama censurados (Efron, 1975)). Esto suele ocurrir en el caso en que se estudie el tiempo de supervivencia desde el diagnóstico de una enfermedad hasta la muerte de los pacientes, ya que no siempre se puede prolongar el estudio hasta que se observa la muerte de todos los sujetos. De esta forma, en la mayor parte de los estudios hay sujetos en los que se ha producido el evento de interés y sujetos censurados.

Se dice que un dato está censurado a la derecha cuando el evento de interés ocurre después de un tiempo preestablecido, este tiempo determina el final del estudio, o cuando el sujeto abandona el estudio por alguna causa. Para estos casos, denotamos como C_i el tiempo desde el comienzo hasta el final del estudio o hasta el abandono del sujeto del estudio, este tiempo se llama tiempo de censura. Se define $T_i^* = \min(T_i,C_i)$ como el tiempo de seguimiento.

Se dice que un dato está censurado a la izquierda cuando se sabe que el evento de interés ha ocurrido para un individuo antes de un instante de tiempo C_i. Para tales individuos sabemos que han experimentado el evento en algún momento antes del tiempo C_i, pero el momento exacto de ocurrencia es desconocido. Este tipo de censura puede ocurrir si algunos individuos experimentan el evento de interés antes de que ocurra el primer periodo de medida. Cuando ocurren censuras a la izquierda los datos del estudio quedan convenientemente representados con un par de variables aleatorias de la forma (T_i^*, δ_i^*). Pero en esta ocasión $T_i^* = \max(T_i,C_i)$.

Otro tipo de censuras, más general, surge cuando se sabe que la ocurrencia del evento de interés sucede dentro de un intervalo de tiempo. A este tipo de censuras se le llama intervalo de censura. Estas censuras se presentan cuando los individuos de un estudio se siguen de forma periódica. De manera que sólo se sabe que el evento de interés ha ocurrido dentro del intervalo delimitado por dos tiempos consecutivos de observación.

Podemos también referirnos a la censura aleatoria, es decir, suponemos que el tiempo de censura C_i para el i-ésimo individuo es una variable aleatoria con función de densidad y función de supervivencia $f_i(s)$ y $S_i(c)$, respectivamente, y además los $C_1, ...,C_N$ son variables aleatorias independientes entre sí.

En nuestro estudio tenemos datos censurados a la derecha, ya que en el registro no se han podido seguir a todos los sujetos hasta la ocurrencia del evento de interés, en esta caso la muerte.

2.3. Riesgos proporcionales

Una asunción clave para poder utilizar el modelo de Cox es que se cumpla la propiedad de riesgos proporcionales. Hay varias formas de probar la propiedad de riesgos proporcionales de los datos (Therneau, 1990). Cuando se trabaja con variables categóricas con un número pequeño de categorías se puede utilizar un test gráfico de la curva de supervivencia. Si se mantienen los riesgos proporcionales, entonces la transformación logarítmica de las curvas de supervivencia de cada categoría debería encontrarse separadas a distancia constante unas de otras. Esta forma de probar la condición de riesgos proporcionales no es útil cuando las variables tienen muchas categorías o cuando las variables son continuas.

Otra posibilidad para estudiar la propiedad de riesgos proporcionales es utilizar los residuales de Schoenfeld (Therneau, 2000). Estos residuales se representan gráficamente, y en el caso de que se cumpla la hipótesis de riesgos proporcionales los residuos deben agruparse de forma aleatoria a ambos lados del valor 0 del eje Y. Para facilitar la interpretación de estos gráficos se suele superponer en la gráfica de residuales una curva de ajuste, que en algunos programas suele ser una curva spline. Cuando se cumple la propiedad de riesgos proporcionales esta curva ajustada debe ser próxima a una línea recta. Los residuos de Schoenfeld son los mas efectivos para detectar anomalías en cada una de las variables que intervienen en el modelo.

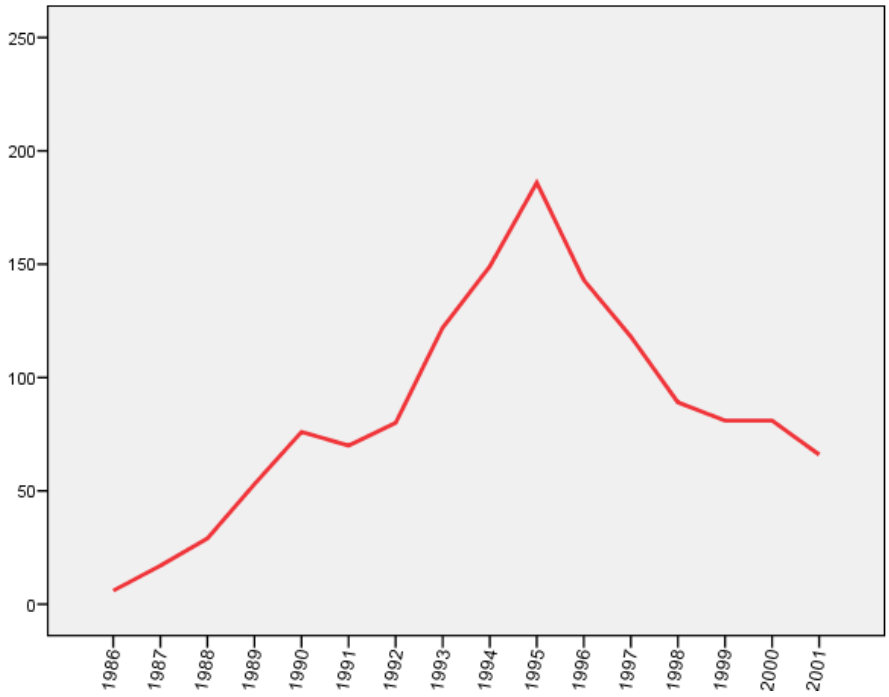
Para la realización de los análisis requeridos en este trabajo se ha utilizado el paquete estadístico S-Plus.

3. SUPERVIVENCIA EN MUJERES CON SIDA EN ANDALUCÍA

3.1. Descripción de los casos

Se ha aplicado el modelo de Cox a datos de mujeres infectadas de SIDA. Se han estudiado 1366 mujeres diagnosticadas en Andalucía entre 1982 y 2001, estos datos han sido proporcionados por el Registro de SIDA de Andalucía (Consejería de Salud, 2006). Este registro forma parte de un sistema nacional de vigilancia epidemiológica en el que participan todas las Comunidades Autónomas de España. El objetivo de la vigilancia del SIDA es conocer su morbilidad, mortalidad y los factores de riesgo asociados a esta enfermedad. Los casos de SIDA son notificados al Registro Andaluz de Casos de Sida a través de los médicos asistenciales, principalmente en el ámbito hospitalario, aunque también se diagnostican casos en atención primaria. Otra institución importante que aporta casos al registro son los centros penitenciarios. Además se hace una búsqueda activa de casos, realizada por técnicos de salud pública, a través de la revisión del CMBD (Conjunto Mínimo Básico de Datos), en el cual se registran las altas hospitalaras de todos los pacientes ingresados en algún hospital andaluz.

Figura 1. Número de mujeres diagnosticadas de SIDA en Andalucía



Las variables disponibles en el registro son: fecha de diagnóstico, fecha de muerte, estado vital, edad, periodo de diagnóstico, provincia de residencia y categoría de transmisión. El tiempo de supervivencia lo hemos definido como el tiempo transcurrido entre la fecha de diagnóstico de la enfermedad y la fecha de muerte, si ésta ha ocurrido antes del día 31 de mayo de 2005. Los individuos que han sobrevivido después de esa fecha se consideran censurados.

En la Figura 1 se muestra la evolución del número de mujeres diagnosticadas de SIDA en Andalucía a lo largo del periodo 1982-2001. El primer caso de SIDA en mujeres se diagnosticó en el año 1986, aumentando el número de casos hasta el año 1995, a partir de este momento comienza un descenso continuado.

Esta disminución del número de casos es consecuencia de la historia natural de la enfermedad, tras el conocimiento de los mecanismos de transmisión de la enfermedad. Además, el descenso ha sido más acusado de lo esperado por la introducción del TARGA en el año 1996.

3.2. Análisis univariante

Mediante el análisis univariante se han representado las curvas de Kaplan-Meier para comprobar si hay diferencias en la supervivencia entre las distintas categoría de cada variable que interviene en el modelo.

Las diferentes curvas de supervivencia las hemos obtenido mediante el estadístico de Kaplan-Meier, que nos da información sobre el valor medio de supervivencia de los sujetos y la mediana del tiempo de supervivencia.

Las tendencias seguidas por los tres grupos de edad durante el periodo de estudio muestran que el grupo de 30-35 años presenta una mayor supervivencia, mientras que el grupo de mas de 35 años es el que presenta una peor supervivencia. (Figura 2).

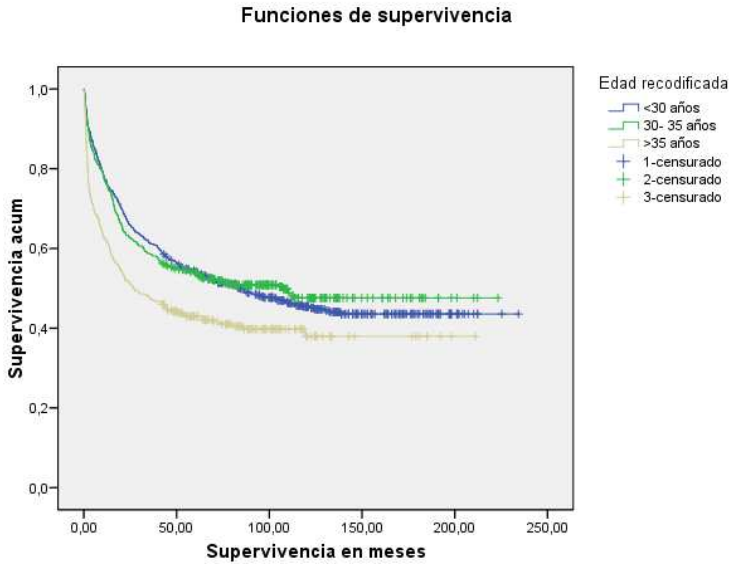


Figura 2. Supervivencia según grupo de edad (Andalucía, 1982-2001)

En la Figura 3 se muestran las tendencias en supervivencia según el periodo de diagnóstico, se ve claramente la mayor supervivencia del periodo 1996-2001, con una supervivencia a los 5 años del diagnóstico del 66%.

En cuanto a la categoría de transmisión, cabe destacar que el grupo de los homosexuales tiene una supervivencia menor que el resto de los grupos hasta aproximadamente los 50 meses desde el diagnóstico, como se puede observar en la Figura 4

Con respecto a la provincia de residencia, Granada y Sevilla son las provincias con mayor supervivencia, con una supervivencia a los 5 años del 67% y 58% respectivamente. La provincia con menor supervivencia es Málaga, con una supervivencia transcurridos 5 años desde el diagnóstico del 44%.

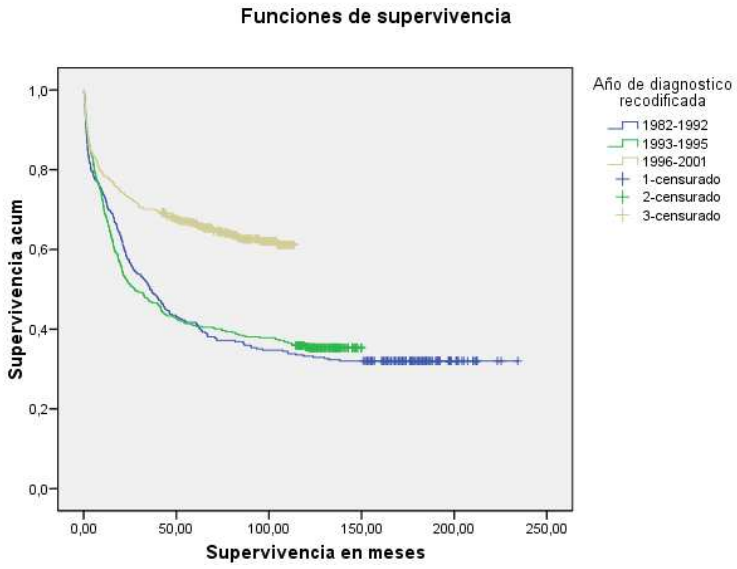


Figura 3. Supervivencia según periodo de diagnóstico (Andalucía, 1982-2001)

.

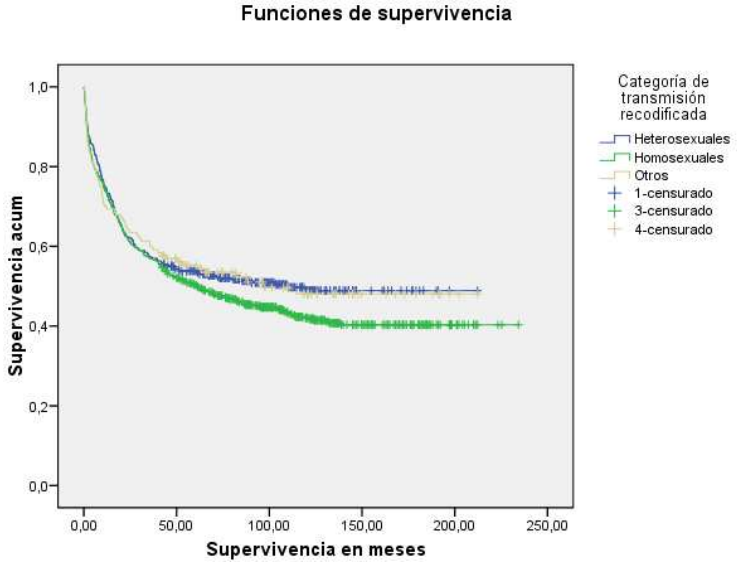


Figura 4. Supervivencia según categoría de transmisión (Andalucía, 1982-2001)

3.3. Análisis multivariante

El modelo que hemos utilizado, para analizar las variables que influyen en la supervivencia de las mujeres diagnosticadas de SIDA en Andalucía, es el modelo de Cox. Para decidir que variables forman parte del modelo hemos utilizado el test de la razón de verosimilitud. El modelo definitivo que explica la supervivencia de las mujeres con SIDA en Andalucía está formado por las variables edad, periodo de diagnóstico y provincia de residencia.

El modelo que explica la supervivencia de las mujeres diagnosticadas de SIDA en Andalucía está formado por las variables edad, periodo de diagnóstico y provincia de residencia. Los resultados del modelo se muestran en la Tabla 1.

Tabla 1. Variables que influyen en la supervivencia de SIDA.

Variable	Coeficiente	e.e.	RR	I.C. RR	p-valor
Edad					
< 30 años	-	-	1	-	-
30-35 años	0.84	0.10	2.3	(1.89-2.82)	<0.001
> 35 años	0.78	0.09	2.2	(1.82-2.62)	<0.001
Periodo de diagnóstico					
1996-2001	-	-	1	-	-
1993-1995	0.59	0.10	1.80	(1.49-2.18)	<0.001
1982-1992	0.13	0.10	1.13	(0.95-1.15)	0.162
Provincia de diagnóstico					
Granada	-	-	1	-	-
Almería	0.35	0.19	1.43	(0.99-2.05)	0.055
Cádiz	0.34	0.16	1.40	(1.03-1.91)	0.032
Córdoba	0.54	0.18	1.71	(1.19-2.45)	0.004
Huelva	0.38	0.20	1.47	(0.99-2.16)	0.054
Jaén	0.31	0.22	1.37	(0.89-2.09)	0.150
Málaga	0.54	0.15	1.72	(1.28-2.32)	<0.001
Sevilla	0.14	0.17	1.15	(0.83-1.60)	0.408

Para estudiar si hay diferencias en la supervivencia entre las tres categorías en que se ha dividido la variable edad hemos utilizado como categoría de referencia el grupo formado por las mujeres menores de 30 años. Este grupo es el que presenta una mejor supervivencia, ya que las mujeres con edades comprendidas entre los 30 y los 35 años tienen un riesgo de morir 2.3 veces superior al de grupo mas joven. El grupo de las mayores de 35 años tiene un riesgo de morir 2.2 veces mayor al del grupo más joven.

Controlando por el resto de variables, se encontraron diferencias estadísticamente significativas en la supervivencia entre los diferentes periodos de diagnóstico. Las mujeres diagnosticadas entre 1982 y 1992 presentan un riesgo de muerte 1.13 veces superior a las diagnosticadas en el periodo 1996-2001, aunque no es estadísticamente significativo. Las diagnosticadas entre 1993 y 1995 también presentan mayor riesgo de morir que las diagnosticadas en el periodo 1992-2001, concretamente un riesgo 1.80 veces superior, su intervalo de confianza es (1.49-2.18). Todo esto pone de manifiesto que la introducción del TARGA mejora la supervivencia de las mujeres diagnosticadas de SIDA.

Para estudiar si hay diferencias en la supervivencia entre las distintas provincias andaluzas se ha considerado a Granada como la categoría de referencia. Controlando por el resto de variables, la mayor parte de las provincias presentan una peor supervivencia que Granada, siendo las residentes en Málaga las que presentan un mayor riesgo de morir, con un riesgo 1.72 veces mayor que las residentes en Granada, el intervalo de confianza es (1.28-2.32). Sevilla presenta un riesgo de morir 1.15 veces mayor que Granada, aunque este riesgo no es estadísticamente significativo (p-valor =0.408).

3.4. Discusión

El resultado más relevante de nuestro estudio ha sido la mejoría en el tiempo de supervivencia tras la introducción del TARGA en el año 1996. Hecho que ya se ha puesto de manifiesto en otros estudios (Lim, 2005; Macrae, 2005) Como ya hemos comentado, la supervivencia en dicho periodo ha sido un 80% superior a la supervivencia de las mujeres diagnosticadas en el periodo 1993-1995.

Además de las variables incluidas en nuestro estudio, en la literatura aparecen otros factores determinantes del tiempo de supervivencia. Los más estudiados son el recuento de CD4 (Zaccarelli, 2005), la carga viral (Jabs, 2005) y el índice de Karfnoski (Shen, 2005).

En nuestro estudio hubiera sido conveniente tener en cuenta el recuento de CD4, la carga viral y el índice de Karnofsky. Sin embargo, al obtener los datos de un registro de casos de Sida es muy difícil tener disponible esta información para todos los pacientes. Hemos trabajado con una fuente de información secundaria, lo que implica que haya información que no se puede recoger e información que sólo está disponible para un grupo pequeño de pacientes.

A pesar de la limitación en la información recogida en los registros de casos de Sida, éstos tienen un papel fundamental en el conocimiento de la epidemia y en la lucha contra esta enfermedad. El objetivo principal de los registros de casos de Sida en conocer la incidencia de esta enfermedad, a la vez que nos permite conocer los mecanismos de transmisión más frecuentes, las características epidemiológicas de la población afectada, y evaluar los cambios a lo largo del tiempo. Los registros de casos de Sida nos permiten realizar estudios con una gran cantidad de pacientes y con un bajo coste económico y de tiempo.

Para futuras investigaciones sería interesante realizar estudios en los que se pudiera tener en cuenta otros factores asociados al tiempo de supervivencia que no se han podido utilizar en este estudio, como son el recuento de CD4, la carga viral o el índice de Karfnosky.

RECEIVED APRIL 2008
REVISED FEBRUARY 2009

REFERENCIAS

[1] COLLET D. (1994): **Modelling survival data in medical research**. Chapman and Hall, London.

[2] CONSEJERÍA DE SALUD. (2006): **Informe anual VIH/SIDA 2.005**. Sevilla

[3] COX DR. (1972): Regression models and life-tables. **J.Roy.Statist.Soc.** B 34, 187-220.

[4] COX, D.R. A and OAKES, D. (1984): **Analysis of Survival Data**. Chapman and Hall, London.

[5] EFRON, B. (1976): The efficiency of Cox´s likelihood function for censored data. **J. Amer. Satist. Assoc**. 72, 557-565.

[6] JABS DA, HOLBROOK JT, VAN NATTA ML, CLARK R, JACOBSON MA, KEMPEN JH ET AL. (2005): Risk factors for mortality in patients with AIDS in the era of highly active antiretroviral therapy. **Ophthalmology** 112, 771-9.

[7] LIM ST, KARIM R, TULPULE A, NATHWANI and BN, LEVINE AM. (2005): Prognostic factors in HIVrelated diffuse large-cell lymphoma: before versus after highly active antiretroviral therapy. **J Clin Oncol**., 23, 8477-82.

[8] MACRAE J, FRIEDMAN AL, EGGERS P, FRIEDMAN EA. (2005): Improved survival in HIV-infected African-Americans with ESRD. **Clin Nephrol.,** 64,124-8.

[9] PACHÓN, J; PUJOL, E; RIVERO, A. (2000): **La infección por el VIH: guía práctica**. Consejería de Salud. Sociedad Andaluza de Enfermedades Infecciosas, Sevilla..

[10] SHEN JM, BLANK A, SELWYN PA. (2005): Predictors of mortality for patients with advanced disease in an HIV palliative care program. **J Acquir Immune Defic Syndr**; 40, 445-7.

[11] THERNEAU, T; GRAMBSCH, P; FLEMING, T. (1990):Martingale based residual for survival models. **Biometrika**, 77, 147-160.

[12] THERNEAU, T; GRAMBSCH, P. (2000): **Modeling Survival Data: Extending the Cox Model**. Springer-Verlag, Berlin..

[13] ZACCARELLI M, TOZZI V, LORENZINI P, TOTTA MP, FORBICI F, VISCO-COMANDINI U ET Al. (2005). Multiple drug class-wide resistance associated with poorer survival after treatment failure in a cohort of HIV-infected patients. **AIDS**;19, 1081-1089.