

# Redes neuronales artificiales aplicadas al análisis de supervivencia: un estudio comparativo con el modelo de regresión de Cox en su aspecto predictivo

Alfonso Palmer Pol y Juan José Montaña Moreno  
Universidad de las Islas Baleares

El objetivo de este estudio fue comparar el rendimiento en predicción entre los modelos de Redes Neuronales Artificiales (RNA) y el modelo de riesgos proporcionales de Cox en el contexto del análisis de supervivencia. Más concretamente, se intentó comprobar: a) si el modelo de redes neuronales jerárquicas es más preciso que el modelo de Cox, y b) si el modelo de redes neuronales secuenciales supone una mejora respecto al modelo de redes neuronales jerárquicas. La precisión fue evaluada a partir de medidas de resolución (área bajo la curva ROC) y calibración (prueba de Hosmer-Lemeshow) usando un conjunto de datos de supervivencia. Los resultados mostraron que las redes neuronales jerárquicas tienen un mejor rendimiento en resolución que el modelo de Cox, mientras que las redes secuenciales no suponen una mejora respecto a las redes neuronales jerárquicas. Finalmente, los modelos de RNA proporcionan curvas de supervivencia más ajustadas a la realidad que el modelo de Cox.

*Artificial neural networks applied to the survival analysis: A comparative study with Cox regression model in its predictive aspect.* The purpose of this study was to compare the performance in prediction between the models of Artificial Neural Networks (ANN) and Cox proportional hazards models in the context of survival analysis. More specifically, we tried to verify: a) if the model of hierarchical neural networks is more accurate than Cox's model, and b) if the model of sequential neural networks signifies an improvement with respect to the hierarchical neural networks model. The accuracy was evaluated through resolution (the area under the ROC curve) and calibration (Hosmer-Lemeshow test) measures using survival data. Results showed that hierarchical neural networks outperform Cox's model in resolution while sequential neural networks do not suppose an improvement with respect to hierarchical neural networks. Finally, ANN models produced survival curves that were better adjusted to reality than Cox's model.

La presencia de información incompleta o censurada constituye una característica fundamental en los datos de supervivencia que hace difícil su manejo mediante los métodos estadísticos convencionales (Allison, 1995). En este tipo de datos también se pueden utilizar variables dependientes del tiempo, esto es, variables cuyos valores pueden cambiar a lo largo del período de observación.

El modelo de regresión de riesgos proporcionales, conocido habitualmente como modelo de regresión de Cox (Cox, 1972), es el modelo más utilizado en este contexto y relaciona la función de riesgo con las variables explicativas por medio de la expresión:

$$h(t, X) = h_0(t)e^{\beta' X}$$

Un aspecto importante del modelo de Cox radica en que éste se puede utilizar para realizar predicciones sobre el proceso de cambio. Más concretamente, en el presente trabajo, nos proponemos utilizar el modelo de Cox para predecir la función de supervivencia, con unos determinados valores en las variables explicativas. La función de supervivencia para un sujeto dado se puede obtener mediante el modelo de Cox a través de la siguiente expresión:

$$S(t, X) = S_0(t)e^{-\beta' X}$$

Para que una variable explicativa pueda formar parte de este modelo se debe verificar si ésta cumple el «supuesto de proporcionalidad» (Allison, 1984). En el caso que se incumpla este supuesto, habitualmente se excluye del modelo la variable explicativa y ésta se trata como variable de estrato (Blossfeld y Rohwer, 1995; Marubini y Valsecchi, 1995; Parmar y Machin, 1995).

La utilización de las Redes Neuronales Artificiales (RNA) se ha centrado principalmente en la clasificación de patrones y en la estimación de variables cuantitativas, sin embargo, apenas existen aplicaciones en el campo del análisis de supervivencia. En este sentido, podemos considerar pioneros los trabajos de Ohno-Ma-

chado (Ohno-Machado, Walker y Musen, 1995; Ohno-Machado y Musen, 1997a; Ohno-Machado y Musen, 1997b), quien ha propuesto dos modelos de red neuronal que permiten el manejo de datos de supervivencia sin necesidad de imponer ningún supuesto de partida, susceptibles de ser un buen complemento al modelo de Cox: el modelo de redes jerárquicas y el modelo de redes secuenciales.

El modelo de redes jerárquicas (Ohno-Machado, Walker y Musen, 1995) consiste en una arquitectura jerárquica de redes neuronales del tipo perceptrón multicapa que predicen la supervivencia mediante un método paso a paso (ver figura 1). De este modo, cada red neuronal se encarga de dar como salida la probabilidad de supervivencia en un intervalo de tiempo determinado, proporcionando el modelo general la supervivencia para el primer intervalo, después para el segundo intervalo y así sucesivamente.

El modelo de redes secuenciales (Ohno-Machado y Musen, 1997a; Ohno-Machado y Musen, 1997b) supone una ampliación respecto al modelo de redes jerárquicas. En el modelo de redes secuenciales la predicción realizada por una red neuronal para un intervalo de tiempo puede actuar a su vez como variable explicativa o de entrada en otra red dedicada a la predicción de otro intervalo anterior o posterior (ver figura 2). Con esta estrategia, se pretende modelar explícitamente la dependencia temporal que existe entre las predicciones realizadas en los diferentes intervalos de tiempo y así obtener curvas de supervivencia asintóticamente decrecientes.

Con el presente estudio se pretende comprobar, por un lado, si el modelo de redes jerárquicas presenta un rendimiento superior en cuanto a predicción frente al modelo de regresión de Cox y, por otro lado, si el modelo de redes secuenciales supone una mejora en rendimiento respecto al modelo de redes jerárquicas. Estas dos hipótesis serán contrastadas a partir de un conjunto de datos de supervivencia derivado del campo de las conductas adictivas.

Materiales y métodos

Matriz de datos

Los datos utilizados en la presente investigación proceden de una serie de estudios realizados por el equipo de McCusker (McCusker et al., 1995; McCusker, Bigelow, Frost et al., 1997; McCusker, Bigelow, Vickers-Lahti, Spotts, Garfield y Frost, 1997) en la Universidad de Massachusetts (la matriz de datos, denominada uis.dat, se puede obtener en la sección *Survival Analysis* de la siguiente dirección URL: <http://www-unix.oit.umass.edu/~statdata>). El objetivo de estos estudios fue comparar diferentes programas de

intervención diseñados para la reducción del abuso de drogas en una muestra de 628 toxicómanos. Estos programas podían diferir en función de la duración de la intervención (corta o larga) y de la orientación terapéutica (clínica A o clínica B). En la tabla 1 se presenta la descripción de las nueve variables explicativas utilizadas en la investigación. No se utilizaron variables dependientes del tiempo. La variable de respuesta fue el tiempo en días transcurrido desde el inicio del estudio hasta la recaída del sujeto en el consumo de drogas. Por tanto, el suceso de interés fue el cambio de estado de no consumo a consumo de drogas. El seguimiento de los sujetos se realizó a lo largo de tres años y medio.

En el gráfico 1 se puede observar la representación gráfica del estimador Kaplan-Meier de la función de supervivencia sobre los datos del estudio.

A partir del valor de los deciles obtenidos mediante Kaplan-Meier, se determinaron diferentes intervalos de tiempo. El decil 9 no fue utilizado debido a que los valores censurados se acumulan al final del seguimiento, como puede observarse en el gráfico, y apenas hay cambios de estado en ese período. De esta forma, se obtuvieron ocho intervalos de tiempo en los que la probabilidad de supervivencia se va decreciendo de forma aproximadamente

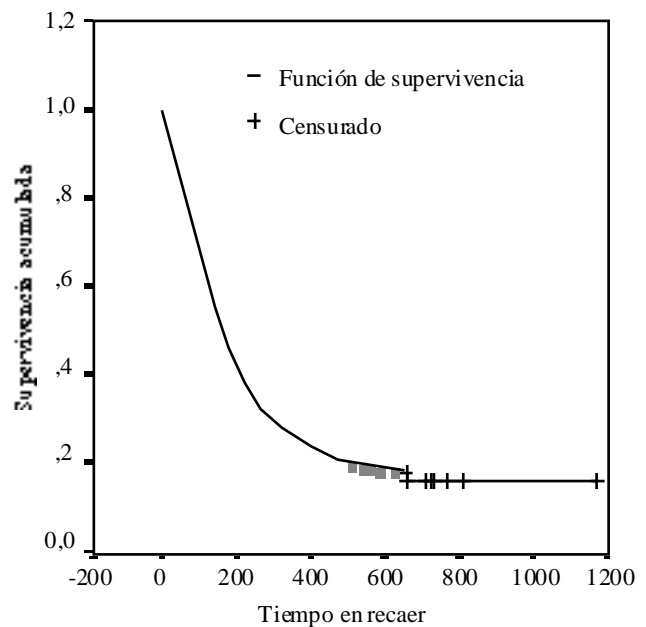


Gráfico 1. Función de supervivencia estimada mediante el método Kaplan-Meier

Tabla 1 Descripción de las variables explicativas		
Variante	Descripción	Valores
Edad	Edad al entrar en el estudio	años
Beck	Puntuación en el Inventario de Depresión de Beck	0 - 54
Hercoc	Uso de heroína/cocaina 3 meses antes de la admisión	1= heroína y cocaína; 2= solo heroína; 3= solo cocaína; 4= ni heroína, ni cocaína
Hdroga	Historia del uso de droga	1= nunca; 2= anterior; 3= reciente
Ntrat	Número de tratamientos previos contra la droga	0 - 40
Raza	Raza del sujeto	0= blanca; 1= no blanca
Trat	Asignación aleatoria de tratamiento	0= corto; 1= largo
Lugar	Lugar del tratamiento	0= clínica A; 1= clínica B
Durac	Duración de estancia en tratamiento (fecha de admisión a fecha de salida)	

constante a medida que avanza el seguimiento. A continuación, se procedió a dividir aleatoriamente la muestra total en dos grupos de forma que la proporción de cambio en cada intervalo era aproximadamente la misma en ambos grupos: 528 sujetos actuaron como grupo de entrenamiento para la construcción de los modelos y 100 sujetos actuaron como grupo de test para la comparación entre modelos. En la tabla 2 se muestra, para cada intervalo de tiempo considerado, los días de seguimiento que comprende el intervalo, la distribución acumulada de cambios y no cambios de estado y la proporción acumulada de cambio, para el grupo de entrenamiento y el de test.

#### Modelo de Cox

Para la generación del modelo de Cox se procedió, en primer lugar, a comprobar el supuesto de proporcionalidad en las nueve variables explicativas. Se pudo observar que la variable «duración de estancia en tratamiento» (Durac) no cumple el supuesto de proporcionalidad y, en consecuencia, quedó excluida del modelo para ser utilizada como variable de estratificación con dos estratos. A continuación, fueron introducidas en el modelo las ocho variables explicativas restantes y todos los términos de interacción de primer orden (los términos de interacción de segundo orden y de orden superior no fueron introducidos, debido a que el método de estimación de los parámetros del modelo no alcanzaba la convergencia). Se generaron variables ficticias para aquellas variables nominales con más de dos categorías. Mediante un método de selección paso a paso hacia atrás basado en la razón de verosimilitud (*backward stepwise: likelihood ratio*), quedaron incluidas en el modelo cuatro variables y siete términos de interacción.

#### Modelos de redes neuronales

El modelo de redes neuronales jerárquicas estaba compuesto por ocho redes del tipo perceptrón multicapa como las presentadas en la figura 1, cada una estaba centrada en dar como salida la probabilidad de supervivencia en uno de los intervalos de tiempo creados.

A fin de obtener el modelo de red óptimo en cuanto a predicción y evitar así el fenómeno del sobreajuste, se utilizó un grupo de 100 sujetos de entrenamiento seleccionado aleatoriamente como grupo de validación. La configuración neuronal que exhibiera el mejor rendimiento ante el grupo de validación sería el modelo seleccionado para pasar a la fase de test. Se probaron diferentes arquitecturas en cuanto al número de neuronas en la capa oculta, funciones de activación de las neuronas y algoritmos de aprendizaje como el *backpropagation* (Rumelhart, Hinton y Williams,

1986) y alguna de sus variantes más utilizadas: *quickpropagation* (Fahlman, 1988), *delta-bar-delta* (Jacobs, 1988), gradiente conjugado (Battiti, 1992), *resilient propagation* (Smith, 1993). Finalmente, se utilizaron ocho redes perceptrón multicapa con dos neuronas en la capa oculta, función de activación tangente hiperbólica en la capa oculta y lineal en la capa de salida, y entrenadas mediante el algoritmo de gradiente conjugado.

En el modelo de redes secuenciales la predicción realizada por una red neuronal para un intervalo de tiempo actúa a su vez como variable explicativa o de entrada en otra red dedicada a la predicción de otro intervalo anterior o posterior. De esta forma, el intervalo correspondiente a la primera red neuronal actúa como intervalo informativo y el intervalo correspondiente a la segunda red neuronal actúa como intervalo informado (figura 2). Siguiendo este esquema se cruzaron los ocho modelos jerárquicos correspondientes a los ocho intervalos de tiempo, generándose 56 redes secuenciales.

Para la generación de las arquitecturas neuronales se empleó el programa *Neural Connection 2.1* (SPSS Inc., 1998) que permite simular el comportamiento de una red perceptrón multicapa asociada al algoritmo de gradiente conjugado.

#### Técnicas de comparación

Se comparó la eficacia de los modelos presentados a partir de las predicciones realizadas sobre los 100 sujetos de test. La eficacia en cuanto a predicción se determinó a partir de medidas de resolución y calibración.

La resolución hace referencia a la capacidad de discriminar por parte del modelo entre sujetos que realizan el cambio de estado y sujetos que no realizan el cambio. La resolución se midió a partir del área bajo la curva ROC (*Receiver Operating Characteristics*) (Swets, 1973, 1988). La comparación entre dos áreas bajo la curva ROC se realizó mediante la prueba z descrita por Hanley y McNeil (1983).

La calibración hace referencia a lo cerca que se encuentran las probabilidades proporcionadas por el modelo respecto al resultado real. La calibración se midió a partir de la prueba  $\chi^2$  de Hosmer-Lemeshow (Hosmer y Lemeshow, 1980).

#### Resultados

##### Comparación modelo de Cox versus modelo de redes jerárquicas

En relación a la comparación en función de la resolución, el gráfico 2 muestra las áreas bajo la curva ROC del modelo de Cox

Tabla 2  
Distribución acumulada de casos en grupo de entrenamiento y test de acuerdo al intervalo de tiempo

Intervalo	Cambio	Grupo de Entrenamiento			Propor.	Cambio	Grupo de Test		
		No Cambio	Total				No Cambio	Total	Propor.
1º 1-26	53	475	528	0.10	10	90	100	0.10	
2º 27-59	106	422	528	0.20	20	80	100	0.20	
3º 60-90	161	367	528	0.30	30	70	100	0.30	
4º 91-120	212	316	528	0.40	39	61	100	0.39	
5º 121-166	265	263	528	0.50	50	50	100	0.50	
6º 167-220	320	208	528	0.60	60	40	100	0.60	
7º 221-290	371	157	528	0.70	70	30	100	0.70	
8º 291-501	423	104	527	0.80	79	21	100	0.79	

y el modelo de redes jerárquicas. Ambos modelos presentan una buena precisión diagnóstica, excepto en el caso del modelo de Cox para el intervalo número ocho que proporciona un área bajo la curva ROC por debajo de 0.70 (Swets, 1988). Se puede observar que

el modelo de redes exhibe un rendimiento superior en todos los intervalos considerados. Esta superioridad se comprueba a nivel estadístico mediante la prueba z (con riesgo unilateral) (Hanley y McNeil, 1983). Por otra parte, se observa que el error estándar del

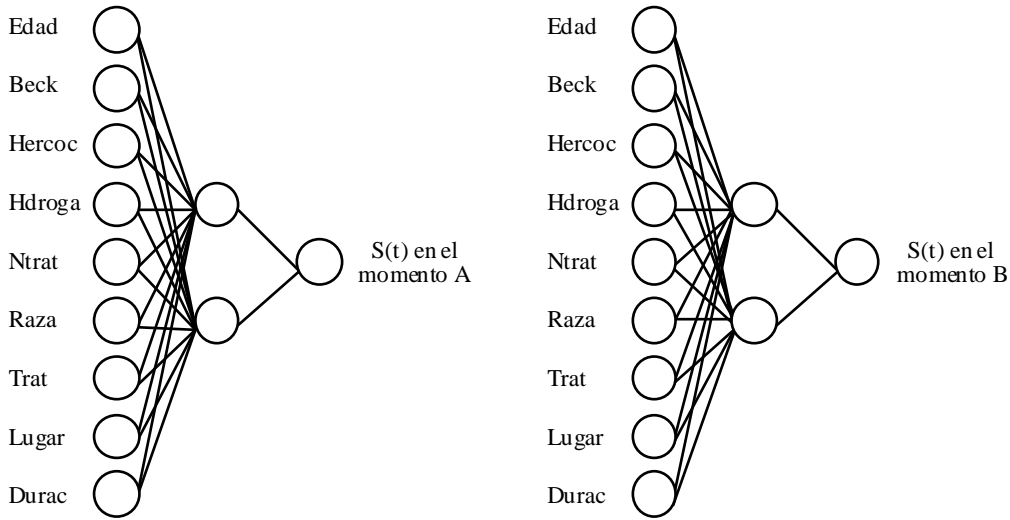


Figura 1. Modelo de red jerárquica

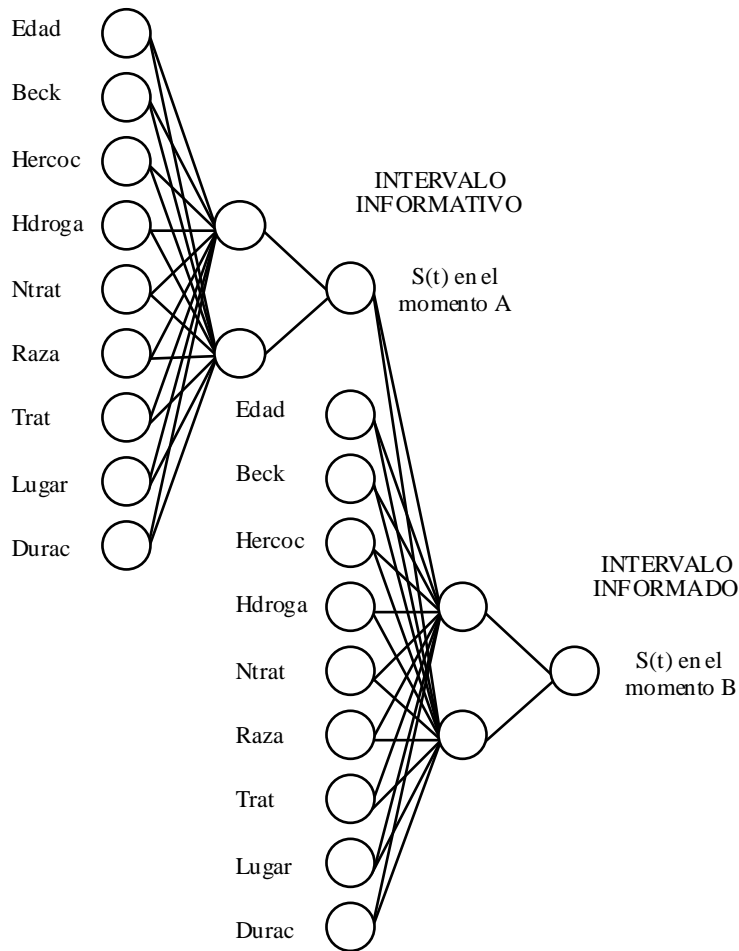


Figura 2. Modelo de red secuencial

área bajo la curva ROC obtenido con el modelo de redes es sistemáticamente inferior que el obtenido con el modelo de Cox. Este hecho implica que las estimaciones del área bajo la curva ROC del modelo de redes tendrán más precisión que en el caso del modelo de Cox.

En relación a la calibración, en la tabla 3 se puede observar a través de la prueba de Hosmer-Lemeshow (1980) que ambos modelos tienen un buen ajuste en todos los intervalos de tiempo considerados, debido a que la discrepancia entre lo observado y lo esperado no es significativa. En este caso, no se aprecian diferencias importantes entre el modelo de Cox y el modelo de redes jerárquicas respecto a la medida de calibración.

*Comparación modelo de redes jerárquicas versus modelo de redes secuenciales*

En relación a la comparación en función de la resolución, la tabla 4 muestra los resultados de la prueba z (Hanley y McNeil, 1983) que permite comparar el área bajo la curva ROC del mo-

delo de redes jerárquicas y del modelo de redes secuenciales. Los valores z positivos indican un mejor rendimiento por parte del modelo de redes secuenciales, mientras que los valores z negativos indican un mejor rendimiento por parte del modelo de redes jerárquicas. En ningún caso, las redes secuenciales mostraron un rendimiento significativamente superior frente a la versión jerárquica. Más bien, se puede observar que el rendimiento de las redes secuenciales fue inferior en numerosos casos. En este sentido, el ejemplo más significativo es la red que predice la probabilidad de supervivencia en el intervalo número cinco utilizando como intervalo informativo el número tres ( $z = -4.76, p < .01$ ).

En relación a la calibración, la tabla 5 muestra los resultados de la prueba de Hosmer-Lemeshow (1980) para las 56 redes secuenciales. Todos las redes secuenciales mostraron un buen ajuste, aunque no supuso una mejora en rendimiento respecto al modelo de redes jerárquicas (ver tabla 5). Se puede apreciar que, en general, el rendimiento mejoraba cuando se utilizaba como intervalo informativo el intervalo inmediatamente posterior.

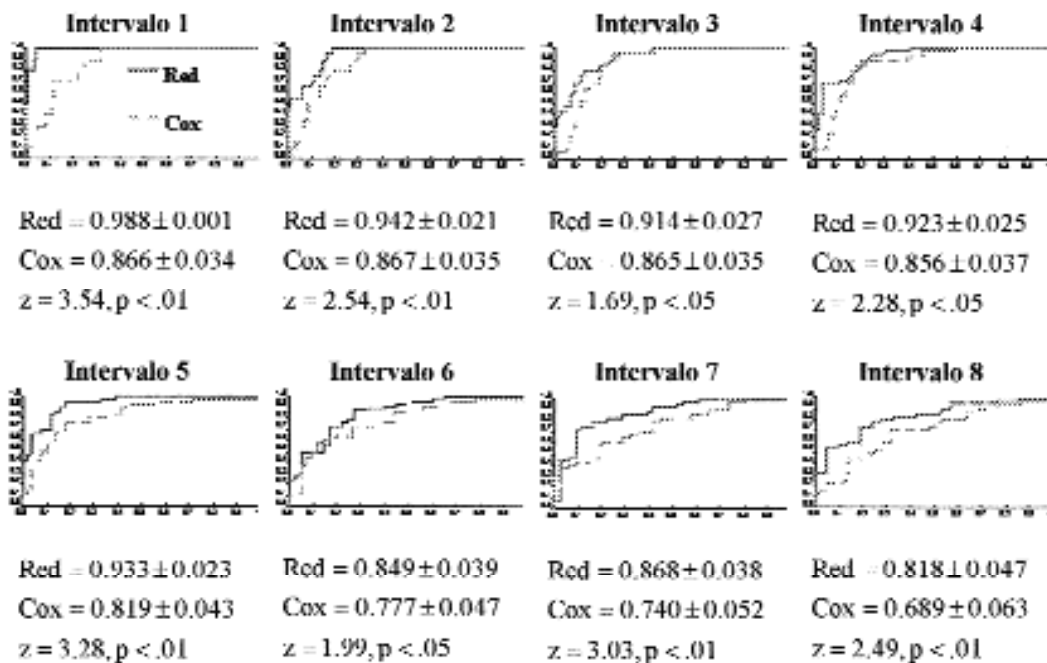


Gráfico 2. Áreas bajo la curva ROC del modelo de redes jerárquicas y el modelo de Cox

Intervalo	Prueba $\chi^2$	
	Modelo de Cox	Redes jerárquicas
1	0.654	1.516
2	1.053	1.928
3	4.373	1.941
4	4.067	4.786
5	4.632	4.469
6	5.574	5.393
7	5.334	5.270
8	5.112	6.050

Con el objeto de demostrar que los modelos de RNA analizados también pueden servir para generar curvas de supervivencia ajustadas tanto para sujetos como para grupos, se muestran en el gráfico 3 las curvas de supervivencia estimadas por los tres modelos –modelo de Cox, modelo de redes jerárquicas y modelo de redes secuenciales generado mediante la utilización del intervalo posterior a cada momento como intervalo informativo– para un sujeto perteneciente al grupo de test que realizó el cambio de estado en el intervalo número cinco. En la gráfica se aprecia cómo los modelos de redes se ajustan más a la realidad que el modelo de Cox. Más concretamente, para el intervalo en que se produce el cambio, las redes jerárquicas y secuenciales proporcionan una es-

*Tabla 4*  
Comparación en resolución entre el modelo de redes jerárquicas y el modelo de redes secuenciales

Prueba z	Intervalo informado							
	1	2	3	4	5	6	7	8
<b>Intervalo informativo</b>								
1		-0.66	0.06	-0.21	-0.07	1.05	-0.41	0.25
2	-0.21		-1.03	0.15	0.01	0.77	0.06	0.33
3	0.11	0.36		-0.32	-4.76**	0.66	-0.46	0.24
4	0.00	0.00	0.70		0.46	1.45	-0.35	-0.78
5	0.00	-0.15	-0.30	0.17		1.08	0.03	0.46
6	0.13	-0.04	0.20	-0.65	-1.04		-0.51	-0.12
7	0.00	0.20	0.46	0.27	-0.38	0.72		-0.10
8	0.13	0.41	0.02	0.05	-0.01	1.12	-0.12	

Nota: \*\* p <.01 prueba unilateral

*Tabla 5*  
Calibración del modelo de redes secuenciales

Prueba $\chi^2$	Intervalo informado							
	1	2	3	4	5	6	7	8
<b>Intervalo informativo</b>								
1		1.620	1.216	0.952	5.281	3.196	7.305	5.982
2	0.859		0.241	2.883	5.799	5.004	3.407	3.544
3	2.478	1.196		2.100	4.164	6.551	9.204	9.404
4	1.570	1.932	0.972		5.075	2.044	5.011	7.035
5	1.484	1.929	1.641	2.950		7.071	5.072	8.068
6	1.646	1.806	0.620	2.018	3.151		2.723	1.897
7	1.668	1.658	1.405	3.959	4.489	1.343		7.793
8	1.481	2.419	0.812	2.581	4.356	4.015	6.767	

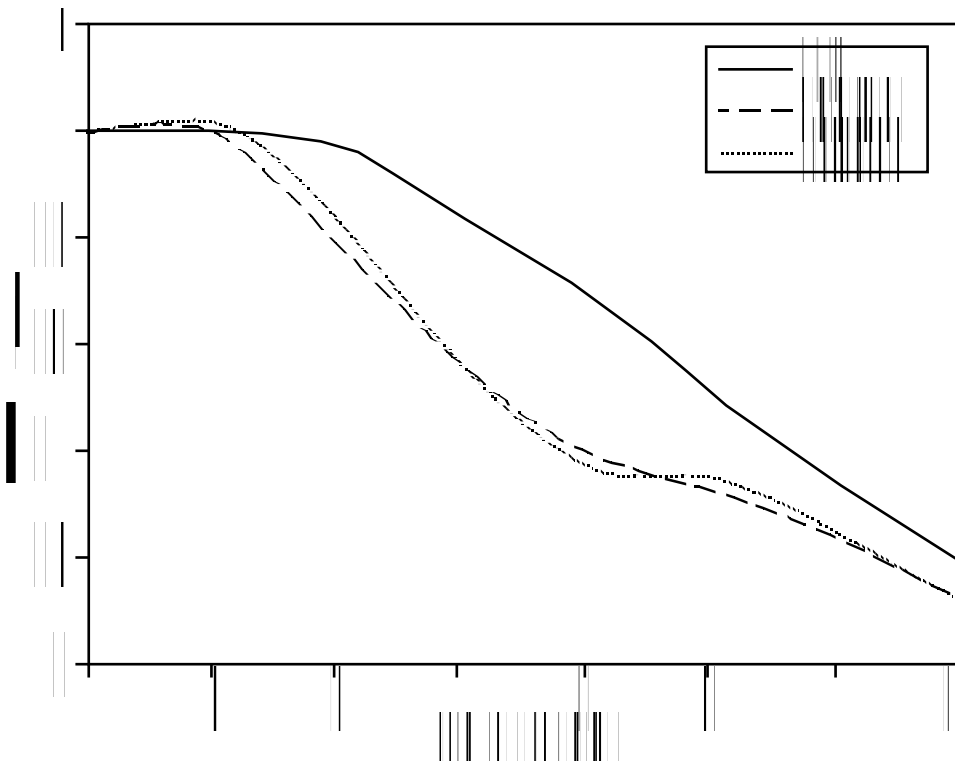


Gráfico 3. Estimación de la función de supervivencia de los tres modelos analizados para un sujeto de test que realizó el cambio en el intervalo 5

timación de la función de supervivencia de 0.3694 y 0.4005, respectivamente; mientras que para el modelo de Cox esta estimación es de 0.6947. Por otra parte, se observa que el modelo de redes secuenciales suaviza ligeramente la curva respecto al modelo de redes jerárquicas. Aunque en algunos casos se han observado pequeñas anomalías en las curvas de supervivencia obtenidas mediante los modelos de redes, para la mayoría de individuos estas curvas son monótonamente decrecientes y, como en el ejemplo comentado, más ajustadas a la realidad que el modelo de Cox.

### Conclusiones

En el presente estudio se ha comprobado que los modelos de RNA jerárquicos y secuenciales permiten el manejo de datos de supervivencia sin necesidad de imponer supuestos de partida en los datos. La información parcial proporcionada por los datos censurados es utilizada en aquellas redes neuronales para las que se tiene información del cambio de estado en el intervalo de tiempo correspondiente. Por ejemplo, los datos de un sujeto al que se le haya realizado el seguimiento hasta el tercer intervalo considerado serán usados en las redes correspondientes al primer, segundo y tercer intervalo, pero no en las redes correspondientes a los siguientes intervalos de tiempo. Si bien en este trabajo no se han utilizado variables dependientes del tiempo, éstas se pueden incorporar fácilmente debido a que cada red neuronal puede recibir, en ca-

da momento temporal, un valor diferente respecto a las variables explicativas para un mismo sujeto.

La comparación llevada a cabo en cuanto a poder predictivo entre los modelos presentados ha puesto de manifiesto, en primer lugar, que el modelo de redes secuenciales tiene una resolución significativamente mejor que el modelo de Cox, mientras que ambos modelos han mostrado una calibración similar. En segundo lugar, el modelo de redes secuenciales no ha supuesto una mejora en rendimiento respecto al modelo de redes jerárquicas, observándose en algunos casos una peor ejecución por parte del primer modelo. Así pues, con los datos manejados en este estudio, no se obtienen las ventajas descritas por Ohno-Machado (1996) en cuanto a las redes secuenciales. En el trabajo de Ohno-Machado (1996), las redes jerárquicas mostraron un rendimiento superior tanto en resolución como calibración frente al modelo de Cox en la mayoría de intervalos de tiempo considerados. Por su parte, las redes secuenciales no mostraron un mejor rendimiento en calibración respecto al modelo de redes jerárquicas, al igual que en el presente trabajo, aunque en la mayoría de intervalos de tiempo sí obtuvieron una mejor resolución.

Por último, se ha comprobado la utilidad de los modelos de red para realizar curvas de supervivencia tanto individuales como grupales, exhibiendo éstas un mejor ajuste en la estimación de la función de supervivencia frente al modelo de Cox.

Este conjunto de resultados pone de manifiesto que las RNA pueden ser útiles en el análisis de datos de supervivencia.

### Referencias

- Allison, P.D. (1984). *Event history analysis. Regression for longitudinal event data*. Beverly Hills, CA: Sage Pub.
- Allison, P.D. (1995). *Survival analysis using the SAS system: a practical guide*. Cary, NC: SAS Institute Inc.
- Battiti, R. (1992). First and second order methods for learning: between steepest descent and Newton's method. *Neural Computation*, 4(2), 141-166.
- Blossfeld, H.P. y Rohwer, G. (1995). *Techniques of event history modeling*. Mahwah, NJ: Lawrence Erlbaum Associates, Pub.
- Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34, 187-202.
- Fahlman, S.E. (1988). Faster-learning variations on back-propagation: an empirical study. En: D. Touretsky, G.E. Hinton y T.J. Sejnowski (Eds.). *Proceedings of the 1988 Connectionist Models Summer School* (pp. 38-51). San Mateo: Morgan Kaufmann.
- Hanley, J.A. y McNeil, B.J. (1983). A method of comparing the areas under receiver operating characteristics curves derived from the same cases. *Radiology*, 148, 839-843.
- Hosmer, D.W. y Lemeshow, S. (1980). A goodness-of-fit test for the multiple logistic regression model. *Communications in Statistics*, A10, 1.043-1.069.
- Jacobs, R.A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks*, 1(4), 295-308.
- Marubini, E. y Valsecchi, M.G. (1995). *Analysing survival data from clinical trials and observational studies*. New York: John Wiley and Sons.
- McCusker, J., Bigelow, C., Frost, R., Garfield, F., Hindin, R., Vickers-Lahti, M. y Lewis, B.F. (1997). The effects of planned duration of residential drug abuse treatment on recovery and HIV risk behavior. *American Journal of Public Health*, 87, 1.637-1.644.
- McCusker, J., Bigelow, C., Vickers-Lahti, M., Spotts, D., Garfield, F. y Frost, R. (1997). Planned duration of residential drug abuse treatment: efficacy versus treatment. *Addiction*, 92, 1.467-1.478.
- McCusker, J., Vickers-Lahti, M., Stoddard, A.M., Hindin, R., Bigelow, C., Garfield, F., Frost, R., Love, C. y Lewis, B.F. (1995). The effectiveness of alternative planned durations of residential drug abuse treatment. *American Journal of Public Health*, 85, 1.426-1.429.
- Ohno-Machado, L. (1996). *Medical applications of artificial neural networks: connectionist models of survival*. Tesis doctoral no publicada. Stanford University.
- Ohno-Machado, L. y Musen, M. (1997a). Modular neural networks for medical prognosis: quantifying the benefits of combining neural networks for survival prediction. *Connection Science: Journal of Neural Computing, Artificial Intelligence and Cognitive Research*, 9(1), 71-86.
- Ohno-Machado, L. y Musen, M. (1997b). Sequential versus standard neural networks for pattern recognition: an example using the domain of coronary heart disease. *Computational Biology in Medicine*, 27(4), 267-281.
- Ohno-Machado, L., Walker, M. y Musen, M. (1995). Hierarchical neural networks for survival analysis. *Medinfo*, 8 Pt 1, 828-832.
- Parmar, M.K.B. y Machin, D. (1995). *Survival analysis: a practical approach*. New York: John Wiley and Sons.
- Rumelhart, D.E., Hinton, G.E. y Williams, R.J. (1986). Learning internal representations by error propagation. En: D.E. Rumelhart y J.L. McClelland (Eds.). *Parallel distributed processing* (pp. 318-362). Cambridge, MA: MIT Press.
- Smith, M. (1993). *Neural networks for statistical modeling*. New York: Van Nostrand Reinhold.
- SPSS Inc. (1998). *Neural Connection 2.1* [Programa para ordenador]. SPSS Inc. (Productor). Chicago: SPSS Inc. (Distribuidor).
- Swets, J.A. (1973). The relative operating characteristic in psychology. *Science*, 182, 990-1.000.
- Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1.285-1.293.