

Redes neuronales artificiales en Medicina Intensiva. Ejemplo de aplicación con las variables del MPM II

J. TRUJILLANO CABELLO^a, M. BADÍA CASTELLÓ^b, J. MARCH LLANES^b, Á. RODRÍGUEZ POZO^a,
L. SERVIÁ GOIXART^a Y A. SORRIBAS TELLO^b

^aServicio de Cuidados Intensivos. Hospital Universitario Arnau de Vilanova. Lleida. España.

^bDepartamento de Ciencias Médicas Básicas. Universidad de Lleida. España.

Objetivo. La aplicación del índice de gravedad *Mortality Probability Model* (MPM II) en nuestra Unidad de Cuidados Intensivos (UCI) nos muestra una mala calibración que nos obliga a buscar modelos reajustados.

Diseño. Para el reajuste de los modelos utilizamos técnicas basadas en regresión logística (RL) y una red neuronal artificial (RN) (perceptrón multicapa con retropropagación del error). Para valorar estos modelos utilizamos un grupo de 964 pacientes que se dividen en un grupo de Desarrollo (736 pacientes) y un grupo de Validación (228). Se calcula el modelo MPM II *Admisión* y a las 24-horas (MPM II-0 y MPM II-24), los modelos reajustados por RL (RLR-0 y RLR-24) y los obtenidos por las RN (RN-0 y RN-24). Los modelos desarrollados se contrastan en el grupo de Validación evaluando sus propiedades de discriminación con el área bajo la curva ROC (ABC [IC 95 %]) y su calibración con el test de Hosmer-Lemeshow C (HLC [p]).

Resultados. Los modelos MPM II-0 y MPM-24 obtienen una buena discriminación (ABC > 0,8) con pobre calibración (HLC > 25). Los modelos reajustados (RLR y RN) mejoran en calibración manteniendo una aceptable discriminación. La RN es mejor en discriminación (ABC = 0,85 [0,79-0,90]) y calibración (HLC = 21 [p = 0,005]) en el modelo 24-horas, pero sin alcanzar significación.

Conclusión. Una RN es capaz de estratificar el riesgo de mortalidad hospitalaria utilizando las

variables del sistema MPM II. En el mismo grupo de pacientes la RN obtiene diferentes probabilidades de muerte. Esto se asocia con una contribución diferente de las variables en los modelos basados en RL o RN.

PALABRAS CLAVE: predicción de resultados, MPM II, redes neuronales artificiales.

ARTIFICIAL NEURONAL NETWORKS IN INTENSIVE MEDICINE. AN EXAMPLE OF APPLICATION WITH MPM II VARIABLES

Objective. The application of Mortality Probability Model (MPM II) index of severity in our ICU shows poor calibration that forces us to seek re-adjusted models.

Design. For the readjustment of the models we have used techniques based on Logistic Regression (LR) and an Artificial Neuronal Network (ANN) (multilayer perceptron with error retropropagation). In order to assess these models we analyzed a group of 964 patients divided into a development subgroup (736 patients) and a validation subgroup (228 patients). We calculated the MPM II model at admission and at 24-hours (MPM II-0 and MPM II-24), the models readjusted by LR (LRR-0 and LRR-24), and the models obtained through the ANN (ANN-0 and ANN-24). Developed models were verified in the validation subgroup for evaluating their discrimination properties through the area under the ROC (receiver operating characteristic) curve (AURC [CI 95%]), and for its calibration with the Hosmer-Lemeshow C (HLC [p]) test.

Results. MPM II-0 and MPM-24 models showed high discrimination (AURC 0.8) with poor calibration (HLC 25). The readjusted models (LRR and ANN) improved in calibration maintaining an ac-

Correspondencia: Dr. J. Trujillano Cabello.
Servicio de Cuidados Intensivos.
Hospital Universitario Arnau de Vilanova.
Avda. Alcalde Rovira Roure, 80.
25198 Lleida.
Correo electrónico: jtruji@cmb.udl.es

Manuscrito aceptado el 4-XI-2004.

ceptable discrimination. ANN is better for discrimination (AURC = 0.85 [0.79-0.90]) and calibration (HLC = 21 [p = 0.005]) in the 24-hour model, however, without significance.

Conclusion. ANN can stratify the risk of hospital mortality using the variables of the MPM II system. ANN shows different death probabilities in the same group of patients. This is associated with a different contribution from the variables in LR- and ANN-based models.

KEY WORDS: results prediction, MPM II, artificial neuronal networks.

INTRODUCCIÓN

Los modelos de predicción de mortalidad hospitalaria se emplean de forma habitual en los enfermos que ingresan en la Unidad de Cuidados Intensivos (UCI). Estos modelos se utilizan para evaluar los resultados de cada UCI y son necesarios, entre otras finalidades, para estratificar correctamente los enfermos a la hora de participar en ensayos clínicos controlados¹. Entre otros, el sistema MPM II (*Mortality Probability Model*) es uno de los modelos más utilizados en nuestro entorno. Este sistema se desarrolló utilizando la metodología de regresión logística (RL) con una gran base de datos que incluía 19.124 pacientes dentro de un amplio estudio europeo y norteamericano (*European/North American Study of Severity Systems*)².

El sistema MPM II consta de dos partes: el modelo *Admisión*, que incluye las variables tomadas en el momento de admisión en UCI, y el modelo *24-horas*, con variables registradas durante las primeras 24 horas de ingreso en la unidad. Este sistema ha sido evaluado en diferentes grupos de pacientes y ha mostrado ser una buena herramienta para la estratificación de riesgo³.

La fiabilidad de un modelo de predicción se mide determinando su habilidad de distinguir los pacientes supervivientes y los que fallecerán (discriminación), y el grado de correspondencia entre la mortalidad observada y la esperada (calibración)⁴. En la práctica clínica, la aplicación de un modelo determinado en una UCI concreta puede mostrarnos una falta de precisión debida, en la mayoría de ocasiones, a una pérdida de calibración. La utilización de modelos específicos para cada UCI se hace necesaria si queremos valorar aspectos de control de calidad⁵. En nuestra UCI, por las características propias que nos alejan de las condiciones medias de la población que sirvió para confeccionar el MPM II, necesitamos adaptar o reajustar este modelo para conseguir una correcta estratificación del riesgo de muerte en nuestros pacientes. Para reajustar este modelo, podemos emplear metodología estadística, basada en RL, o también utilizar como alternativa un modelo basado en redes neuronales artificiales (RN).

Las RN constituyen una alternativa a los modelos basados en técnicas estadísticas. Las RN aportan una

metodología general como modelos de predicción de resultados que son función de las correspondientes variables de entrada. Mediante un proceso de aprendizaje (llamado también entrenamiento) las RN desarrollan la capacidad de predicción de la variable resultado, en nuestro caso la probabilidad de mortalidad hospitalaria, a partir de un grupo de casos suficientemente amplio. Una vez entrenada correctamente la RN adquiere la capacidad de generalización: es capaz de enfrentarse a casos nuevos (no utilizados en la fase de aprendizaje) y dar resultados correctos⁶. Según los resultados publicados, las RN son entre un 5%-10 % más precisas que otras técnicas estadísticas, ya que no trabajan con restricciones de modelos lineales y son capaces de incorporar automáticamente las relaciones entre las variables predictoras⁷.

En Medicina Intensiva las RN se han utilizado en diversos problemas. En la tabla 1 mostramos algunos de los trabajos más representativos con sus características fundamentales⁸⁻²⁵.

Los objetivos de nuestro trabajo son en primer lugar validar el modelo MPM II en nuestra UCI y compararlo con un modelo construido reajustando los coeficientes correspondientes a las variables de la ecuación logística, y en segundo construir un modelo basado en la metodología de RN con las mismas variables. Los dos modelos (RL y RN) se validan en un grupo independiente de pacientes (cronológicamente posterior). Comparar el modelo MPM II, no sólo con un modelo reajustado, sino con un modelo basado en RN, constituye la contribución de nuestro estudio.

PACIENTES Y MÉTODO

Pacientes

El estudio se ha realizado en la UCI polivalente (médica y quirúrgica) de 10 camas ubicada en el Hospital Universitario Arnau de Vilanova de Lleida. Se informó al Comité Ético del Hospital de la realización del estudio, no se precisó consentimiento informado, ya que todas las variables se recogían para el diagnóstico y el tratamiento de los pacientes y se aseguró en todo momento el anonimato de los mismos.

Se recogieron, de forma prospectiva, los ingresos de enero de 1996 a diciembre de 1999 que formaron el grupo de Desarrollo, y de enero de 2000 hasta diciembre de 2000 que constituyeron el grupo de Validación. Todos los datos demográficos y las variables necesarias para calcular los modelos MPM II-0 (*Admisión*) y MPM II-24 (*24-horas*) fueron recogidos por un equipo entrenado².

Modelo MPM II reajustado según regresión logística

Por las características de nuestra serie (tamaño y porcentaje de mortalidad) aplicamos la estrategia de reajustar los coeficientes de la RL de acuerdo con el método sugerido por Zhu et al³.

TABLA 1. Aplicaciones de las redes neuronales artificiales en Medicina Intensiva

	Referencia	Año	Valida.	Estad.	Result.	Nº
Scores						
APACHE II	Wong	1999	Sí	CS	=	8.796
APACHE III	Clermont	2001	Sí	CS	=	1.647
TRISS-ISS	DiRusso	2000	Sí	CS	+	10.609
ICISS-TRISS-ISS	Rutledge	1998	Sí	CS	+	7.276
Pronóstico						
Sepsis	Warner	1996	Sí	No	+	68
SIRS-Shock	Dybowski	1996	Sí	Sí	++	258
TCE	Lang	1997	Sí	Sí	=	1.066
Cirugía cardíaca	Orr	1997	Sí	Sí	=	1.112
Bypass Ao-Cor	Lippmann	1997	Sí	Sí	=	80.606
Hemorragia cerebral	Edwards	1999	Sí	Sí	++	81
Trasplante hepático	Doyle	1994	No	No	++	155
Weaning	Ashutosh	1992	Sí	Sí	=	25
Post RCP	Ebell	1993	No	No	+	218
Estancia UCI						
> 15 días	Buchman	1994	Sí	Sí	+	491
Cirugía cardíaca	Tu	1993	Sí	No	++	1.409
Diagnóstico						
IAM	Selker	1995	Sí	Sí	+	5.773
Hipomagnesemia	Verive	2000	Sí	Sí	+	463
Niveles aminoglucósidos	Yamamura	1998	Sí	Si	++	30

Referencia: primer autor; Valida.: utilización grupo de validación; Estad.: comparación con técnica estadística; CS: comparación con *score*; Result.: resultado global; =: similar; +: ligeramente superior la RN; ++: mejor la RN; Nº: número de pacientes en el trabajo; TCE: traumatismo craneoencefálico; Bypass Ao-Cor: Bypass aorto-coronario; RCP: reanimación cardiopulmonar; IAM: infarto agudo de miocardio.

Reajustar un modelo de RL supone calcular los nuevos coeficientes de las variables predictoras utilizando los valores propios de cada UCI²⁶.

Diseño de la red neuronal artificial

Para el desarrollo de nuestra aplicación utilizamos una arquitectura de perceptrón multicapa (MLP-*multilayered perceptron*) entrenado con algoritmo de retropropagación del error (*backpropagation*). Para introducirse en el tema existen revisiones sencillas en la literatura^{6,27}.

Una RN (fig. 1) consta de capas que están formadas por nodos. La capa de entrada recibe los valores de las variables predictoras, la capa oculta es donde

se realizan los procesos de ajuste, y la capa de salida donde se obtienen los resultados de la red (*output*). Los nodos están unidos por unos parámetros internos (pesos = *weights*) cuyos valores se modifican de acuerdo a los valores que reciben los nodos y según la función de red que se les aplica. La complejidad de una RN depende de la cantidad de parámetros internos que contenga.

La RN precisa de un proceso de entrenamiento para alcanzar la capacidad de generalización, primero comenzamos con valores aleatorios de los pesos que deben de modificarse de forma iterativa (cada vez que presentamos los datos de entrada a la red) para conseguir los resultados más próximos a los valores observados. La finalidad es minimizar el error

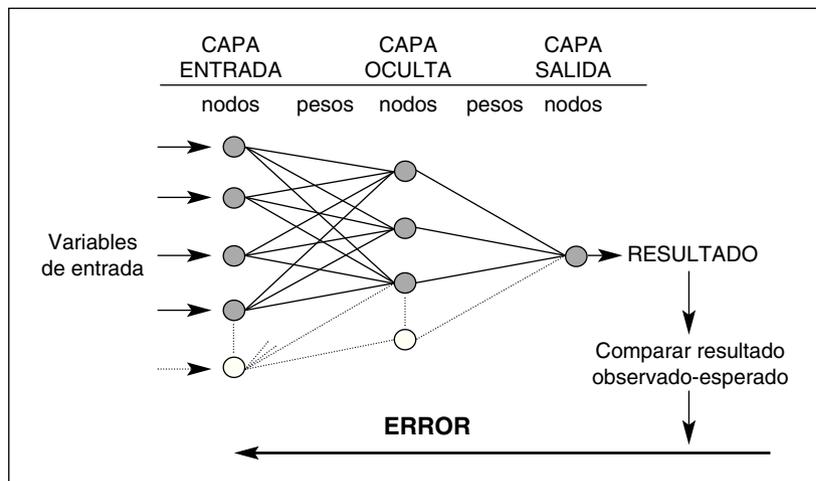


Figura 1. Arquitectura de una red neuronal artificial (perceptrón de tres capas). Los pesos son los parámetros que deben ajustarse durante el proceso de entrenamiento para conseguir que la capa de salida de resultados coincida con los observados. El error entre lo observado y el resultado de la red se propaga hacia atrás (*backpropagation*) y debe ir disminuyendo en las sucesivas iteraciones en las que se presentan los valores de las variables predictoras. La complejidad de una red depende del número de nodos de su capa oculta.

que se observa de esta diferencia entre los valores de la red y los observados.

La estrategia que se sigue para evitar que la red sobreaprenda (sólo memoriza, pero no tiene capacidad de generalización) es dividir el grupo de Desarrollo en un subgrupo de Entrenamiento (70% del grupo de desarrollo) y un subgrupo de Verificación (30% restante). El subgrupo de Verificación no se emplea en el ajuste de los parámetros, pero la red debe mantener buenos resultados en este subgrupo, durante el entrenamiento, para asegurar que la red adquiere la capacidad de generalización.

La arquitectura óptima de la red (número de capas ocultas y de nodos que las forman) se alcanza con un método empírico observando la evolución del error en el subgrupo de Verificación²⁸. El valor de salida de nuestra red es la probabilidad de muerte y se interpreta con un valor de 0 a 1. Para la creación de la RN utilizamos el programa comercial Qnet 97 (*Vesta Services Inc.*). Los parámetros de entrenamiento modificables en el *software* (momento, coeficiente de aprendizaje, etc.) fueron optimizados para alcanzar el mejor resultado de la red.

Comparación de los modelos

Para comparar los distintos modelos medimos su capacidad de discriminación y calibración; la discriminación calculando el porcentaje de pacientes correctamente clasificados (PCC) con el punto de corte a 0,5 y con el área bajo la curva ROC (ABC)^{29,30}; y la calibración usando el test de Hosmer-Lemeshow C³¹, construyendo la curva de calibración y calculando las razones de mortalidad estandarizada (RME)³². Estos cálculos se realizan tanto en el grupo de Desarrollo como en el de Validación. Utilizamos el test de Bland-Altman para analizar las probabilidades generadas por cada modelo³³ y el coeficiente de correlación intraclase (CCI) para valorar la concordancia entre los mismos³⁴. El análisis estadístico se realizó con el programa SPSS (versión 11.0).

RESULTADOS

Características del grupo de estudio

Se analizan un total de 964 pacientes divididos en el grupo de Desarrollo (736) y el de Validación (228).

Nuestra población se define por tener pocos pacientes programados (quirúrgicos), que nos conduce a una mortalidad hospitalaria alta y un incremento en la estancia media en UCI. No encontramos diferencias significativas en los distintos subgrupos definidos en el estudio (tabla 2). Tampoco apreciamos diferencias significativas al analizar las características demográficas en los distintos años que abarcaba el estudio.

Modelo MPM II reajustado según regresión logística

Al reajustar los coeficientes del MPM II se obtienen nuevos modelos: el reajustado en el modelo de

TABLA 2. Características demográficas del grupo de estudio (964 pacientes)

	Desarrollo (n = 736)	Validación (n = 228)	p ^c
Edad (años) ^a	55,9 (18)	57,9 (19)	NS
Sexo, varón (%)	67,5	60,5	NS
Status			NS
Médico (%)	70,7	67,5	
Quirúrgico (%)	23,5	24,1	
Programado (%)	5,8	8,3	
Categoría diagnóstica			NS
Respiratorio (%)	25,8	24,1	
Trauma (%)	25,0	21,5	
Gastrointestinal (%)	14,5	16,2	
Sepsis (%)	7,2	3,1	
Otros (%)	27,4	35,1	
VM (%)	61,3	61,8	NS
MORT (%)	38,0	34,2	NS
Estancia (días) ^b	5 (2-9)	5 (2-11)	NS

Desarrollo: años 1996-1999; Validación: año 2000; VM: ventilación mecánica; MORT: mortalidad hospitalaria; ^amedia (DE); ^bmediana (rango intercuartil) ^cdeterminada por prueba χ^2 para comparación de proporciones, prueba "t" para la comparación de medias y prueba de *Mann-Whitney* para la comparación de medianas; NS: no significativo (p > 0,05).

Admisión (reajuste de regresión logística 0) (RLR-0) y en el modelo de *24-horas* (reajuste de regresión logística 24) (RLR-24).

En el modelo *Admisión*, las variables que no alcanzaron significación estadística fueron: cirrosis, arritmia cardíaca y enfermedad cerebrovascular; en el modelo *24-horas*: cirrosis y tiempo de protrombina mayor de 3 segundos sobre el estándar (tabla 3).

Justificamos las diferencias encontradas en los coeficientes por nuestro diferente *case-mix*, al tener pocos pacientes con cirrosis (sólo 7 pacientes), enfermedad cerebrovascular (33 pacientes) y no incluir a los pacientes coronarios.

Arquitectura de la red neuronal artificial

Las arquitecturas óptimas fueron diferentes para el modelo *Admisión* (con 4 nodos en la capa oculta) que para el modelo *24-horas* (8 nodos). Con redes más complejas (con más nodos en la capa oculta) no mejorábamos la precisión.

Comparación de los modelos

En el grupo de Desarrollo, el modelo MPM II tanto en *Admisión* como a las *24-horas* obtuvo una buena discriminación (ABC de 0,8), pero pobre calibración. Como era esperable, los modelos reajustados (RLR) mejoraron en calibración. Las RN consiguen una mejor discriminación y calibración. En el grupo de Validación se observa similar comportamiento con el modelo MPM II, mejoría en la calibración con los modelos reajustados así como con las redes (tabla 4).

En la figura 2 se muestran las curvas ROC y de calibración del modelo *24-horas* en el grupo de Validación (en el modelo *Admisión* los resultados son simi-

TABLA 3. Variables del modelo MPM II. Coeficientes del modelo original y del reajustado para nuestra UCI según regresión logística

	MPM II		RLR	
	Coeficientes	Odds ratio	Coeficientes	Odds ratio
<i>Admisión</i>				
Constante	-5,468		-5,307	
Coma	1,485	4,4 (3,8-5,2)	0,673	1,9 (1,2-3,7)
FC = 159 x´	0,456	1,6 (1,2-2,1)	1,272	3,5 (1,5-8,3)
TAS ≤ 90 mmHg	1,061	2,9 (2,5-3,4)	0,475	1,6 (1,1-2,5)
Insuficiencia renal crónica	0,919	2,5 (2,0-3,1)	1,344	3,8 (1,8-9,6)
Cirrosis	1,136	3,1 (2,4-4,0)	NS	
Neoplasia metastásica	1,199	3,3 (2,7-4,0)	1,286	3,6 (1,8-17,6)
Insuficiencia renal aguda	1,482	4,4 (3,7-5,2)	1,321	3,7 (2,3-6,0)
Arritmia cardíaca	0,280	1,3 (1,2-1,5)	NS	
ACVA	0,213	1,2 (1,0-1,5)	NS	
Sangrado GI	0,396	1,5 (1,2-1,8)	1,321	3,0 (1,1-9,0)
Masa intracraneal	0,865	2,4 (2,0-2,8)	1,109	9,7 (2,5-13,7)
Edad, 10 años	0,030	1,4 (1,3-1,4)	0,040	1,2 (1,1-1,3)
RCP	0,569	1,8 (1,4-2,2)	0,284	1,3 (1,1-3,7)
VM	0,791	2,2 (2,0-2,5)	0,912	2,4 (1,6-3,7)
Cirugía urgente	1,190	3,3 (2,8-3,8)	1,385	3,9 (1,1-14,4)
<i>24-horas</i>				
Constante	-5,645		-5,205	
Edad, 10 años	0,032	1,4 (1,3-1,4)	0,033	1,1 (1,0-1,1)
Cirrosis	1,087	3,0 (2,3-3,9)	NS	
Masa intracraneal	0,913	2,5 (2,1-3,0)	3,076	11,1 (2,3-18,1)
Neoplasia metastásica	1,161	3,2 (2,6-4,0)	1,545	4,6 (1,1-19,3)
Cirugía urgente	0,834	2,3 (2,0-2,7)	1,017	2,7 (1,1-10,3)
Coma	1,687	5,4 (4,6-6,4)	1,172	3,2 (1,6-6,2)
Creatinina > 2	0,722	2,1 (1,8-2,4)	0,525	1,6 (1,1-2,9)
Infección conf.	0,497	1,6 (1,4-1,9)	0,711	2,0 (1,3-3,0)
VM	0,808	1,7 (1,5-2,1)	1,198	3,3 (2,1-5,1)
PO ₂ < 60 mmHg	0,466	1,6 (1,4-1,9)	0,593	1,8 (1,2-3,9)
T. protromb > 3 seg	0,553	1,7 (1,5-2,1)	NS	
Diuresis < 150 h	0,822	2,3 (1,9-2,7)	0,920	2,5 (1,2-5,0)
Inotropos ≥ 1 h	0,716	2,0 (1,8-2,3)	0,739	2,0 (1,3-3,2)

RLR: modelo de regresión logística reajustado para nuestra UCI; FC: frecuencia cardíaca; TAS: presión arterial sistólica; ACVA: accidente cerebrovascular agudo; GI: gastrointestinal; RCP: reanimación cardiopulmonar; VM: ventilación mecánica; NS: variable no significativa. Definición de variables según artículo original².

lares). Al igual que en las curvas ROC, cuando evaluamos el PCC de los distintos modelos encontramos pocas diferencias, siendo del 74 % en el modelo

MPM II-24, 75 % en RLR-24 y del 77 % en el RN-24 (siendo la diferencia máxima entre los modelos de 4 pacientes correctamente clasificados).

TABLA 4. Comparación de la precisión de los modelos en los grupos de Desarrollo y Validación

	ABC (95 %)	HL-C (p)	RME (95 %)
<i>Desarrollo</i>			
<i>Admisión</i>			
MPM II-0	0,80 (0,77-0,83)	97,4 (0,000)	1,43 (1,32-1,53)
RLR-0	0,82 (0,79-0,85)	15,2 (0,054)	1,00 (0,91-1,08)
RN-0	0,85 (0,82-0,88)*	20,5 (0,008)	0,99 (0,92-1,07)
<i>24-horas</i>			
MPM II-24	0,83 (0,82-0,88)	31,7 (0,000)	1,48 (1,30-1,59)
RLR-24	0,84 (0,81-0,87)	7,0 (0,466)	1,00 (0,92-1,07)
RN-24	0,89 (0,86-0,91)*	14,4 (0,070)	0,99 (0,92-1,06)
<i>Validación</i>			
<i>Admisión</i>			
MPM II-0	0,81 (0,75-0,87)	25,9 (0,001)	1,36 (1,17-1,55)
RLR-0	0,81 (0,75-0,87)	7,4 (0,491)	0,93 (0,78-1,07)
RN-0	0,81 (0,75-0,87)	19,7 (0,011)	0,92 (0,78-1,06)
<i>24-horas</i>			
MPM II-24	0,83 (0,77-0,89)	94,5 (0,000)	1,46 (1,26-1,67)
RLR-24	0,84 (0,78-0,89)	24,3 (0,002)	0,91 (0,78-1,08)
RN-24	0,85 (0,79-0,90)	21,5 (0,005)	0,94 (0,80-1,04)

Desarrollo: años 1996-1999; Validación: año 2000; ABC: área bajo la curva ROC; HL-C: test de Hosmer-Lemeshow C (p > 0,05 se considera un modelo bien calibrado); RME: razón de mortalidad estandarizada; RLR: modelos de regresión logística reajustados; RN: modelo de red neuronal; *ABC es mayor en los modelos RN de forma significativa, pero sólo en el grupo de Desarrollo.

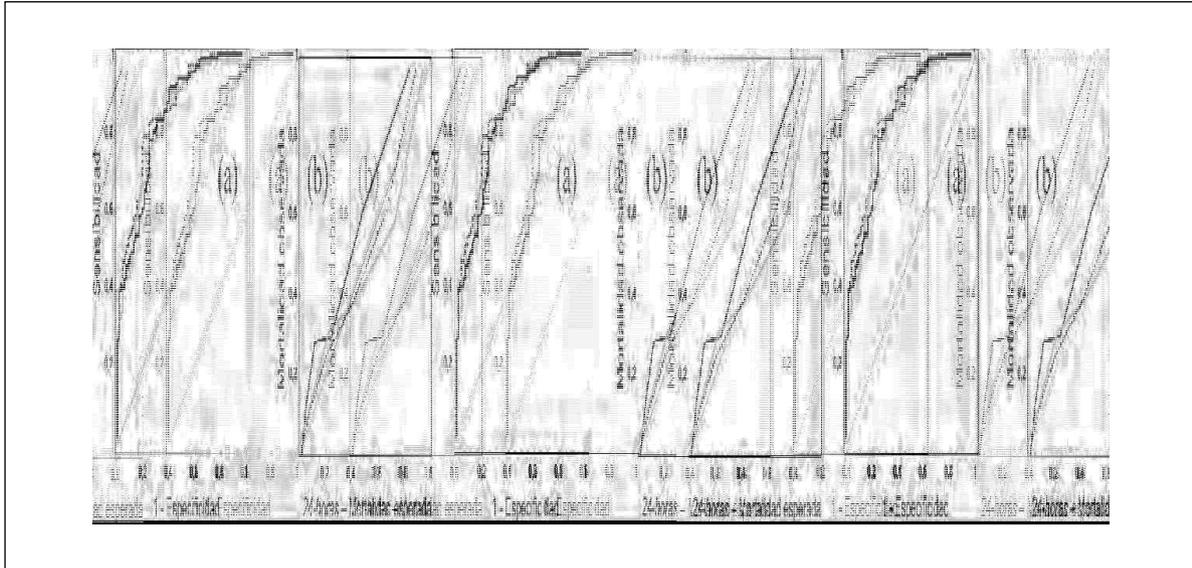


Figura 2. Curvas ROC (a) y curvas de calibración (b) del modelo 24-horas. Línea negra — modelo MPM II, línea gris — modelo de Regresión logística reajustado, y línea de puntos modelo de red neuronal.

En la figura correspondiente a las curvas de calibración observamos la tendencia del modelo MPM II-24 a infraestimar la probabilidad de mortalidad hospitalaria (curva sobre la línea de perfecta calibración).

En el modelo 24-horas, la comparación de las probabilidades calculadas por los modelos mostraban algunas diferencias. Si comparamos el RLR-24 y el MPM-24 (fig. 3a) se observa el efecto conseguido con el reajuste del modelo que consigue una mejor calibración y que muestra un CCI de 0,86 (buena concordancia). La comparación entre el RLR-24 y la RN-24 (fig. 3b) nos da un resultado que no sigue un patrón uniforme, con un CCI que en este caso baja a 0,74.

Los valores de probabilidad de muerte de algunos pacientes son muy diferentes según el modelo utilizado, lo que indicaba la diferente contribución de las variables predictoras según el modelo considerado.

Como ejemplo para explicar diferencias entre los modelos hicimos un estudio de los casos extremos (un total de 136 pacientes), con diferencias de mortalidad superiores en un 20 %, que nos demostró que en algunos casos (en 70 pacientes) pequeñas variaciones en la variable “edad” (no discriminadas por el modelo MPM II) modificaban la probabilidad esperada según la RN, aunque se mantuvieran los mismos valores en el resto de variables predictoras.

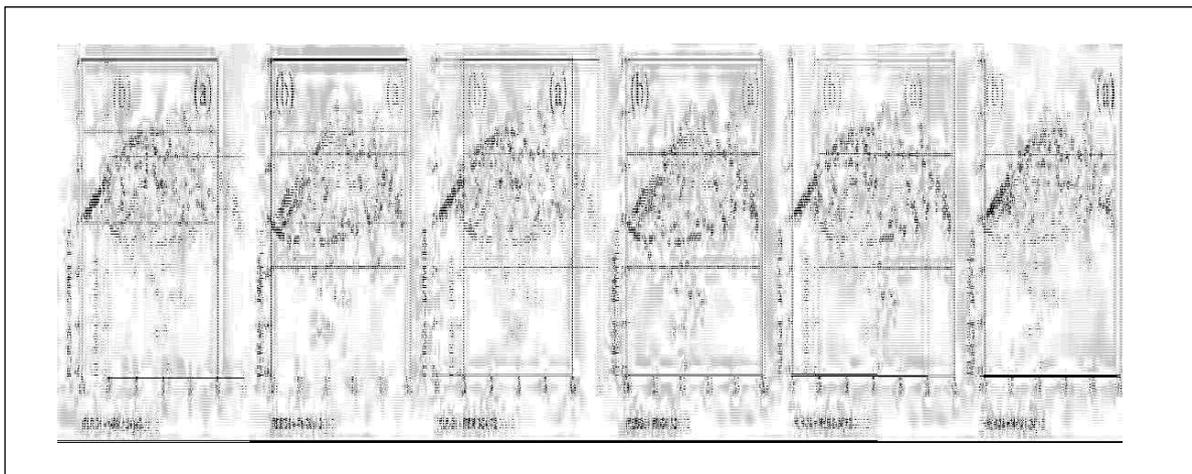


Figura 3. Gráficos de Bland-Altman de los modelos de predicción de mortalidad a las 24-horas en el grupo de Validación. RLR-24: modelo de regresión logística reajustado. RN-24: modelo de red neuronal. Las líneas horizontales corresponden a ± 2 desviaciones estándar. (a): RLR-24 comparado con MPM II-24. (b): RLR-24 comparado con RN-24.

Para evaluar la consistencia de nuestros resultados comprobamos que la utilización de 50 redes entrenadas con distinto grupo de Verificación, asignado al azar, conseguían resultados similares a los expuestos.

DISCUSIÓN

La aplicación de un modelo de predicción de mortalidad puede hacerse con confianza cuando ha sido validado en la población donde se aplica. Las características de nuestra UCI, por su tamaño, bajo número de ingresos programados, alta mortalidad y *case-mix*, nos apartan de las condiciones medias de la población que sirvió para confeccionar el modelo MPM II. Cuando aplicamos este sistema a nuestra UCI, descubrimos una falta de calibración al obtener una mortalidad superior a la esperada según el modelo. Esta discrepancia nos obliga a evaluar nuestro nivel asistencial y encontrar el modo de obtener un método preciso para estratificar la gravedad de nuestros pacientes que nos sirva como herramienta y control de calidad⁵. Para conseguir reajustar el modelo podemos utilizar la metodología estadística basada en RL o utilizar una RN.

Las RN han sido utilizadas de forma creciente en Medicina, existiendo múltiples revisiones publicadas. Las aplicaciones más frecuentes se sitúan en el diagnóstico clínico, farmacología y predicción de resultados³⁵⁻³⁷.

Nuestros resultados no encuentran diferencias significativas entre el modelo reajustado por RL y el conseguido con RN, aunque apreciamos una tendencia de mejores parámetros de discriminación y calibración con las RN. En diversos artículos con hallazgos similares los autores concluyen que la relación existente entre las variables predictoras es independiente y prácticamente lineal³⁸. Nosotros aportamos la visión de que, aunque de forma global los resultados son similares, encontramos pacientes que tienen diferente asignación de probabilidad según el modelo RL o RN. El análisis de este tipo de pacientes descubre que algunas variables (en nuestro caso la edad) contribuyen de forma diferente para ambos modelos. Así, queda planteada la pregunta de cuál de los dos modelos (RL o RN) es más preciso al utilizarlo en nuestra UCI.

En general, una RN es más precisa cuando se cumplen diversas condiciones: la variable final viene expresada como función compleja de las variables predictoras o cuando existen influencias entre las variables predictoras difíciles de encontrar. Por otra parte, las RN son más complicadas, y sus coeficientes no tienen una interpretación fácil de expresar con el lenguaje habitual³⁹.

En nuestros resultados (sobre todo en el modelo *24-horas*) la metodología basada en RN apunta mejores resultados, ya que es capaz de encontrar relaciones y efectos no lineales entre las variables del MPM II y la probabilidad de muerte.

En la revisión de Sargent, que analiza 28 trabajos sobre predicción de resultados con RN (comparada con RL), encuentra que las RN son equivalentes o su-

periores cuando se analizan series con suficiente número de pacientes, aunque también advierte que debe valorarse el sesgo de publicación (tienden a publicarse los trabajos cuando las RN son por lo menos similares o mejores)⁴⁰.

Las variables empleadas en el modelo MPM II han sido seleccionadas bajo técnicas de RL, lo que también penaliza al modelo RN. Si la selección se hubiera hecho con RN, las variables podrían haber sido otras.

Las limitaciones de nuestro trabajo se basan en dos aspectos. Primero, hemos utilizado un tipo de RN concreto (el perceptrón multicapa con algoritmo de retropropagación del error), utilizar otro tipo de red podría conseguir unos mejores resultados¹⁵; además, el empirismo para obtener la arquitectura óptima siempre debe tenerse en cuenta. Y segundo, nuestra serie es pequeña (debido a las características de nuestra UCI y hospital), lo que limita las conclusiones generales del trabajo. Trabajar con un rango de 5 años también introduce factores de confusión que deben tenerse en cuenta. Otra limitación importante es el aspecto de computación de los datos, ya que para el cálculo de la precisión de las redes debemos hacer procedimientos basados en remuestreo intensivo que no hemos mostrado en el artículo para garantizar una mayor claridad en su lectura.

Nuestros resultados son similares a los publicados comparando el modelo APACHE II y APACHE III con las RN^{8,9}.

En conclusión, nuestro trabajo pretende afirmar la necesidad de validar los instrumentos de predicción en cada UCI y la necesidad posterior de reajustar los modelos (al encontrar una falta de calibración) que puede hacerse con técnicas basadas en RL o RN. Poder acercar, de forma sencilla, la metodología basada en redes también puede considerarse una aportación.

Idealmente, la búsqueda de modelos más precisos debe basarse en que estos sean capaces de mantener su precisión cuando se apliquen en poblaciones concretas. Las técnicas de RN pueden ayudarnos en este objetivo.

BIBLIOGRAFÍA

1. Lemeshow S, Le Gall JR. Modeling the severity of illness of ICU patients. *JAMA* 1994;272:1049-55.
2. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993; 270:2478-86.
3. Zhu BP, Lemeshow S, Hosmer DW, Klar J, Avrunin JS, Teres D. Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization: A simulation study. *Crit Care Med* 1996;24:57-63.
4. Patel PA, Grant BJB. Application of mortality prediction systems to individual intensive care units. *Intensive Care Med* 1999;25:977-82.
5. Zimmerman JE, Wagner DP. Prognostic systems in intensive care: How do you interpret an observed mortality that is higher than expected? *Crit Care Med* 2000;28:258-60.
6. Trujillano J, March J, Sorribas A. Aproximación metodológica al uso de redes neuronales artificiales para la predicción de resultados en medicina. *Med Clin (Barc)* 2004;122 Suppl 1:59-67.

7. Levine RF. Conference concluding remarks. *Cancer* 2001; 91:1696-7.
8. Wong LSS, Young JD. A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural network. *Anaesthesia* 1999;54:1048-54.
9. Clermont G, Angus DC, DiRusso SM, Griffin M, Linde-Zwirble WT. Predicting hospital mortality for patients in the intensive care unit: A comparison of artificial neural networks with logistic regression models. *Crit Care Med* 2001;29:291-6.
10. DiRusso SM, Sullivan T, Holly C, Nealon S, Savino J. An artificial neural network as a model for prediction of survival in trauma patients: Validation for a regional trauma area. *J Trauma* 2000;49:212-23.
11. Rutledge R, Osler T, Emery S, Kromhout-Schiro S. The end of the injury severity score (ISS) and the trauma and injury severity score (TRISS): ICISS, an international classification of diseases, ninth revision-based prediction tool, outperforms both ISS and TRISS as predictor of trauma patient survival. *J Trauma* 1998;44:41-9.
12. Warner A, Bencosme A, Polycarpou MM, Healy D, Verme C, Conway JY, et al. Multiparameter models for the prediction of sepsis outcome. *Ann Clin Lab Science* 1996;26:471-9.
13. Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. *Lancet* 1996;347:1146-50.
14. Lang EW, Pitts LH, Damron SL, Rutledge R. Outcome after severe head injury: an analysis of prediction based upon comparison of neural network versus logistic regression analysis. *Neuro Res* 1997;19:274-80.
15. Orr RK. Use of a probabilistic neural network to estimate the risk of mortality after cardiac surgery. *Med Decis Making* 1997;17:178-85.
16. Lippmann RP, Shahian DM. Coronary artery bypass risk prediction using neural networks. *Ann Thorac Surg* 1997;63:1635-43.
17. Edwards DF, Hollingsworth H, Zazulia AR, Diringer MN. Artificial neural networks improve the prediction of mortality in intracerebral haemorrhage. *Neurology* 1999;53:351-7.
18. Doyle HR, Dvorchik I, Mitchell S, Marino IR, Ebert FH, McMichael J, et al. Predicting outcomes after liver transplantation. A connectionist approach. *Ann Surg* 1994;219:408-15.
19. Ashutosh K, Lee H, Mohan CK, Ranka S, Mehrotra K, Alexander C. Prediction criteria for successful weaning from respiratory support: statistical and connectionist analyses. *Crit Care Med* 1992;20:1295-301.
20. Ebell HM. Artificial neural networks for predicting failure to survive following in-hospital cardiopulmonary resuscitation. *J Fam Pract* 1993;36:297-303.
21. Buchman TG, Kubos KL, Seidler AJ, Siegforth MJ. A comparison of statistical and connectionist models for the prediction of chronicity in a surgical intensive care unit. *Crit Care Med* 1994;22:750-62.
22. Tu JV, Guerriere MRJ. Use of a neural network as a predictive instrument for length of stay in the intensive care unit following cardiac surgery. *Comput Biomed Res* 1993;26:220-9.
23. Selker HP, Griffith JL, Patil S, Long WJ, D'Agostino RB. A comparison of performance of mathematical predictive methods for medical diagnosis: identifying acute cardiac ischemia among emergency department patients. *J Investig Med* 1995; 43:468-76.
24. Verive MJ, Irazuzta J, Steinhart CM, Orłowski JP, Jaimovich DG. Evaluating the frequency rate of hypomagnesemia in critically ill paediatric patients by using multiple regression analysis and a computer-based neural network. *Crit Care Med* 2000;28:3534-9.
25. Yamamura S, Nishizawa K, Hirano M, Momose Y, Kimura A. Prediction of plasma levels of aminoglycoside antibiotic in patients with severe illness by means of an artificial neural network simulator. *J Pharm Sci* 1998;1:95-101.
26. Moreno R, Apolone G, Reis Miranda D. Evaluation of the uniformity of fit of general outcome prediction models. *Intensive Care Med* 1998;24:40-7.
27. Rosenberg AL. Recent innovations in intensive care unit risk-prediction models. *Curr Opin Crit Care* 2002;8(4):321-30.
28. Tourassi GD, Floyd CE, Coleman RE. Acute pulmonary embolism: cost-effectiveness analysis of the effect of artificial neural networks on patient care. *Radiology* 1998;206:81-8.
29. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
30. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839-43.
31. Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;115:92-106.
32. Rapoport J, Teres D, Lemeshow S, Gehlbach S. A method for assessing the clinical performance and cost-effectiveness of intensive care units: A multicenter inception cohort study. *Crit Care Med* 1994;22:1385-91.
33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10.
34. Prieto L, Lamarca R, Casado A. La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclass. *Med Clin (Barc)* 1998;110:142-5.
35. Cross BS, Harrison RF, Kennedy RL. Introduction to neural networks. *Lancet* 1995;346:1075-9.
36. Baxt WG. Application of artificial neural networks to clinical medicine. *Lancet* 1995;346:1135-8.
37. Dayhoff JE, DeLeo JM. Artificial neural networks. Opening the black box. *Cancer* 2001;91:1615-35.
38. Duh MS, Walker AM, Pagano M, Kronlund K. Prediction and cross-validation of neural networks versus logistic regression: using hepatic disorders as an example. *Am J Epidemiol* 1998;147:407-13.
39. Armoni A. Use of neural networks in medical diagnosis. *MD Computing* 1998;15:100-4.
40. Sargent DJ. Comparison of artificial neural networks with other statistical approaches. Results from medical data sets. *Cancer* 2001;91:1636-42.