

Originales

Evaluación de la reproducibilidad de la recogida de datos para el APACHE II, APACHE III adaptado para España y SAPS II en 9 Unidades de Cuidados Intensivos en España

L. DOMÍNGUEZ^a, P. ENRÍQUEZ^a, P. ÁLVAREZ^b, M. DE FRUTOS^c, V. SAGREDO^d, J. LÓPEZ-MESSA^e, D. CARRIEDO^f, F. TABOADA^g, Á. GARCÍA-LABATTUT^h, F. GANDÍAⁱ, M. VALLEDOR^j Y J. BLANCO^a

^aUnidad de Cuidados intensivos. Hospital Universitario Río Hortega. Valladolid.

^bDepartamento de Estadística Aplicada. Universidad de Valladolid.

^cUnidad de Cuidados Intensivos. Hospital General Yagüe. Burgos.

^dUnidad de Cuidados Intensivos. Hospital Clínico de Salamanca.

^eUnidad de Cuidados Intensivos. Hospital Río Carrión. Palencia.

^fUnidad de Cuidados Intensivos. Complejo Hospitalario de León.

^gUnidad de Cuidados Intensivos. Hospital Central de Asturias. Oviedo.

^hUnidad de Cuidados Intensivos. Hospital General de Soria.

ⁱUnidad de Cuidados Intensivos. Hospital Clínico Universitario de Valladolid.

^jUnidad de Cuidados Intensivos. Hospital San Agustín. Avilés.

Objetivo. Evaluar la reproducibilidad en la recogida de datos y su influencia en el cálculo de la gravedad y del riesgo predicho de muerte para los modelos APACHE II, APACHE III adaptado para España y SAPS II.

Diseño. Estudio multicéntrico, prospectivo y observacional de cohortes.

Ámbito. Nueve Unidades de Cuidados Intensivos (UCI) en España.

Pacientes. Inclusión consecutiva de los pacientes ingresados en el período de estudio. Se excluyeron los pacientes menores de 16 años, con estancia en UCI menor de 24 horas, los ingresados para implante programado de marcapasos y los reingresados en UCI dentro del mismo ingreso hospitalario.

Intervención. Ninguna.

Variables de interés principales. Se recogieron los datos necesarios para el cálculo de las puntuaciones de gravedad y del riesgo predicho de muerte. Se seleccionaron el 10% de los pacientes por muestreo aleatorio simple y se recogieron los mismos datos por un grupo independiente de intensivistas. Finalmente se compararon los datos recogidos por los dos grupos de intensivistas.

Resultados. Se encontraron diferencias significativas en el APS (*acute physiology score*) y puntuación de gravedad calculados para el APACHE III y SAPS II, y en el riesgo de muerte predicho por SAPS II. El porcentaje de acuerdos en el diagnóstico de ingreso en UCI fue del 50% para los modelos APACHE II y III. En la mayoría de los pacientes (76,58% en el APACHE II y 79,82% en el APACHE III) la diferencia en el riesgo predicho de muerte debido a la diferente asignación del diagnóstico de ingreso en UCI fue menor del 10%.

Conclusiones. En este estudio el APS se mostró como el factor más influyente en la reproducibilidad de los índices de gravedad y del cálculo del riesgo predicho de muerte. El diagnóstico de ingreso en UCI no mostró un impacto importante en la reproducibilidad del riesgo predicho de muerte.

PALABRAS CLAVE: APACHE, reproducibilidad, cuidados intensivos, predicción de mortalidad y evaluación de resultados.

Correspondencia: Dr. L. Domínguez.
Unidad de Cuidados intensivos.
Hospital Universitario Río Hortega.
C/ Cardenal Torquemada.
47010 Valladolid.
Correo electrónico: ldominguez@hurh.sacyl.es.

Manuscrito aceptado el 3-IV-2007.

EVALUATION OF THE REPRODUCIBILITY OF THE DATA COLLECTION FOR THE APACHE II, APACHE III ADAPTED FOR SPAIN AND THE SAPS II IN NINE INTENSIVE CARE UNITS IN SPAIN

Objective. To assess reproducibility in data collection and its influence on the calculation of the severity scoring and mortality risk in APACHE II, APACHE III adapted for Spain and SAPS II.

Design. Multicenter, prospective, observational cohort study.

Setting. Nine Spanish Intensive Care Units (ICUs).

Patients. 1,211 consecutive patients admitted during the study period were included. Those patients under 16 years of age, those with a stay in the ICU of less than 24 hours, those admitted for scheduled pacemaker implant and those readmitted to the ICU within the same hospital admission were excluded.

Intervention. None.

Endpoints of interest. The data needed to calculate the severity and mortality risk scores were collected. A total of 10% of the patients were chosen by simple random sampling and the same data were collected by an independent group of intensive care physicians. Finally, the data obtained by the two groups of intensivists were compared.

Results. Significant differences were detected in the acute physiology score (APS) and severity score used for the calculation of APACHE III and SAPS II, and the predicted risk of death calculated for SAPS II. The percentage of agreement on admission diagnosis to the ICU was 50% for both APACHE II and III models. Nonetheless, in most of the patients (76.58% for APACHE II and 79.82% for APACHE III), the difference in the predicted risk of death due to the different assignation of diagnoses on admission to the ICU was less than 10%.

Conclusions. In this study, APS was the most influential factor on the reproducibility of severity scores and risk of death prediction. Admission diagnosis assignment had no significant impact on the reproducibility of the predicted mortality risk.

KEY WORDS: APACHE, reliability, intensive care, mortality prediction and outcome process assessment.

INTRODUCCIÓN

Los modelos pronósticos en cuidados intensivos fueron desarrollados para predecir el resultado hospitalario del paciente ingresado en Unidades de Cuidados Intensivos (UCI). Dichos modelos se basan en la medición precoz de la gravedad en UCI, la edad, el estado crónico de salud y el diagnóstico. A pesar de su uso extendido existen varios problemas que limitan su aplicación¹⁴. Por ejemplo, el diferente *case mix* de los pacientes ingresados en una determinada UCI y el de los pacientes en los que el modelo fue desarrollado, la organización de una determinada UCI y del hospital (localización geográfica, lugar donde se

estabiliza al paciente previo al ingreso en UCI, frecuencia de las determinaciones de laboratorio, etc.)⁵⁻⁷.

La reproducibilidad de los datos es un punto clave en lo que concierne a la validez y precisión de los índices de gravedad y del riesgo de muerte predicho por éstos. Son varios los problemas que atañen a la reproducibilidad: unos tienen que ver con los datos (definición de las variables originales, transcripción y conversión de las unidades en el cálculo de variables derivadas y ubicación del paciente con una determinada patología)⁸⁻¹¹ y otros tienen que ver con la variabilidad intra e interobservador¹².

El objetivo del presente estudio es evaluar la reproducibilidad en la recogida de datos y su influencia en el cálculo de las puntuaciones de gravedad y del riesgo predicho de muerte de tres de los modelos pronósticos más utilizados en UCI. Estos modelos son el APACHE II¹³, APACHE III adaptado para España¹⁴ y el SAPS II¹⁵.

PACIENTES Y MÉTODOS

Estudio prospectivo, multicéntrico, observacional de cohortes, realizado en 9 UCI médico-quirúrgicas españolas. Se trata de un subestudio que forma parte de un proyecto para evaluar el funcionamiento de distintos modelos pronósticos en estas Unidades.

Se ingresaron 1.784 pacientes entre noviembre de 1999 y marzo del 2000. Se excluyeron los pacientes menores de 16 años, los ingresados en UCI durante menos de 24 horas, los que ingresaron en UCI para implante de marcapasos definitivo y aquellos reingresados en UCI durante el mismo ingreso hospitalario. Finalmente se analizaron 1.211 pacientes.

Después de un período de entrenamiento de un mes, los datos de los 1.211 pacientes incluidos en el estudio fueron recogidos prospectivamente por residentes e intensivistas pertenecientes a las distintas Unidades (grupo 1). Esta información se obtuvo de la historia clínica del paciente. Posteriormente, se adjudicó un diagnóstico de ingreso por cada paciente y modelo pronóstico siguiendo las especificaciones de los modelos originales¹³⁻¹⁵. Se recogió también la ubicación del paciente previa al ingreso en UCI, las comorbilidades y el estado vital del paciente al alta de la UCI y del hospital. Los datos se introdujeron en una base de datos. Para evitar errores en cuanto a las variables fisiológicas y de laboratorio se tuvo en cuenta el mínimo y máximo valor de cada variable. El programa fue diseñado para escoger aquel valor que peor puntuación otorgaba en cada modelo. Los cálculos del *acute physiology score* (APS), puntuaciones de gravedad y riesgo predicho de muerte se hicieron automáticamente por el programa. Para comprobar el algoritmo de cálculo que seguía el programa, durante 15 días se comprobaron los cálculos efectuados por éste y los cálculos realizados manualmente.

Una vez completada la base de datos se seleccionaron por muestreo aleatorio simple el 10% de los pacientes de cada hospital. En dichos pacientes se realizaron los mismos cálculos por un grupo independiente de intensivistas (grupo 2). En este caso los datos se

ANEXO 1

En los estudios de concordancia para variables cualitativas lo habitual es calcular la proporción de concordantes. Esto quiere decir que sólo se esperan dos resultados posibles (1 si concuerdan, 0 si son discordantes). Sin embargo, esta situación no tiene en cuenta la diferencia real entre los diagnósticos. ¿Son igual de diferentes una «Insuficiencia respiratoria» y un «trauma múltiple» que un «infarto agudo de miocardio» y una «insuficiencia cardíaca» a la hora de evaluar la reproducibilidad del riesgo predicho de muerte? Esto nos muestra la limitación de la metodología habitual a la hora de evaluar la concordancia. Hay algunas discordancias mayores (primera situación) y más importantes, y otras menores (segundo ejemplo), que incluso pueden deberse a que los grupos diagnósticos se superpongan o no estén perfectamente delimitados.

A falta de una medida de la disimilaridad entre los distintos diagnósticos que pudiera haber sido establecida de forma subjetiva por algún grupo de expertos, debemos buscar una medida de lo similares o disimilares que son dos diagnósticos cualesquiera. En nuestro caso, teniendo en cuenta que este estudio de concordancia está orientado a valorar la influencia de ésta en los métodos de predicción del riesgo de muerte, nos hemos basado en estos últimos para dar una medida de la distancia entre diagnósticos.

Cualquier método de predicción del riesgo de muerte funciona de la misma forma. Dados unos valores *input* (diagnóstico de la enfermedad, valores de las variables fisiológicas, etc.) el método proporciona un valor *output* que representa el riesgo de que el paciente fallezca. Cada método en el que es necesario determinar el grupo diagnóstico para obtener un riesgo, asocia –de forma implícita– una curva a cada grupo diagnóstico. Dicha curva se obtiene si variamos todos los demás valores *input*. En el caso del modelo APACHE, para cada grupo diagnóstico tenemos una curva del tipo $R(x, VD) = \exp(-x'b + VD)/(1 + \exp(-x'b + VD))$; donde *VD* es un valor que depende del grupo diagnóstico y *x* representa los valores *input* de las demás variables que influyen. En realidad esta curva determina una medida de probabilidad; es una función de distribución. Esto sugiere utilizar alguna métrica probabilística como medida de lo diferentes o parecidos que son dos diagnósticos. Entre éstas, las métricas de Wasserstein^{16,17} se cuentan entre las más utilizadas, seguramente por sus buenas propiedades. Dadas dos curvas $F(x)$ y $G(x)$ la distancia L_1 de Wasserstein se define como:

$$W(F, G) = \int_0^1 |F(x) - G(x)| dx$$

En nuestro estudio, al tener la misma forma funcional las diferentes curvas asociadas a los diferentes grupos diagnósticos, el cálculo de las distancias es fácil. Dados dos grupos diagnósticos *A* y *B*, cuyo valor diagnóstico en la ecuación APACHE es VD_A y VD_B respectivamente, y si $F_A(x) = R(x, VD_A)$ y $F_B(x) = R(x, VD_B)$ son las respectivas curvas, la distancia entre ellas es: $W_1(F_A, F_B) = |VD_A - VD_B|$. Sin embargo, los valores que proporciona esta distancia no tienen sentido por sí mismos. Para interpretarlos debemos mirar su impacto sobre el riesgo de fallecimiento. Así, si dos diagnósticos distan *h* se puede ver que la máxima diferencia en el riesgo de muerte que asignan es:

$$MDR = \frac{e^{h/2} - e^{-h/2}}{2 + e^{h/2} + e^{-h/2}}$$

Las tablas 6 y 7 muestran el resultado de aplicar esta metodología en nuestro estudio. En cada fila se calcula el número de pacientes cuya distancia entre diagnósticos es igual o inferior al valor de *h* que aparece en la segunda columna. La columna *DR* se utiliza para ver qué diferencia máxima de riesgo de muerte existe entre dos diagnósticos que distan como máximo *h*. De esta forma, para los grupos diagnósticos del APACHE II, el 76,6% de los pacientes muestra un «alto grado de concordancia», pues la máxima diferencia en el riesgo de muerte es del 10%. En el caso del APACHE III es el 78,1% de los pacientes los que muestran este «alto grado de concordancia».

La importancia y potencia de esta metodología radica en su generalidad, ya que sirve también para métodos de predicción cualesquiera, incluyendo aquellos que proporcionan curvas no paramétricas.

recogieron de la historia clínica de forma retrospectiva. Se utilizó el mismo programa informático para realizar los cálculos de las puntuaciones de gravedad y riesgo predicho de muerte. Finalmente, se compararon los datos y cálculos efectuados por el grupo 1 y el grupo 2 de intensivistas.

Para evaluar la calibración de los tres modelos se realizó el test H y C de Hosmer-Lemeshow y se calculó la razón estandarizada de mortalidad (REM). Para evaluar la discriminación se utilizaron las correspondientes curvas ROC.

Para comparar las variables numéricas se utilizaron el coeficiente de correlación intraclass (CCI) y el «t»-test para datos apareados. Para las variables categóricas se calculó el porcentaje de acuerdos. Estas variables fueron: la ubicación del paciente previa al ingreso en UCI (APACHE III), las comorbilidades y el motivo de ingreso en UCI. Se utilizó la métrica probabilística de Wasserstein^{16,17} para calcular la distancia entre los diagnósticos asignados por el grupo 1 y 2 de intensivistas, así como para la evaluación del

impacto de estas distancias sobre el cálculo del riesgo predicho de muerte. Esta metodología comprende dos puntos diferentes:

1. El cálculo de un valor (*h*), que es la diferencia absoluta del peso diagnóstico asignado por cada grupo de intensivistas según el modelo APACHE II y el III.

2. El cálculo de la diferencia de riesgo predicho de muerte (*DR*) debido a la diferente asignación del diagnóstico al mismo paciente por parte de los dos grupos de intensivistas, para el APACHE II y el III. Este método asume que el resto de variables del modelo de predicción son perfectamente reproducibles entre los dos grupos de intensivistas (anexo 1).

Los cálculos estadísticos se realizaron con el programa SPSS 10.0.

RESULTADOS

Las características generales de los 1.211 pacientes se presentan en la tabla 1. La edad media fue de 61,8 años (DE 16,5), la mediana de la estancia en UCI

DOMÍNGUEZ L ET AL. EVALUACIÓN DE LA REPRODUCIBILIDAD DE LA RECOGIDA DE DATOS PARA EL APACHE II, APACHE III ADAPTADO PARA ESPAÑA Y SAPS II EN 9 UNIDADES DE CUIDADOS INTENSIVOS EN ESPAÑA

TABLA 1. Características generales

	Número de pacientes = 1.211
Edad (años)*	61,8 (16,5)
Estancia (días)**	3,06 (2,06; 8,28)
Tipo de paciente	
Médico: número de pacientes (% del total)	773 (63,8)
Quirúrgico: número de pacientes (% del total)	324 (26,8)
Traumatológico: número de pacientes (% del total)	114 (9,4)
Mortalidad en UCI: número de pacientes (% del total)	190 (15,7)
Mortalidad hospitalaria: número de pacientes (% del total)	262 (21,6)

*Media (desviación estándar); **mediana (p₂₅, p₇₅); UCI: Unidad de Cuidados Intensivos.

TABLA 2. Test de Hosmer – Lemeshow y área bajo la curva ROC para los tres modelos

	Test C de Hosmer - Lemeshow	p	Test H de Hosmer - Lemeshow	p	Área bajo la curva ROC (IC 95%)
APACHE II	18,43	< 0,025	18,30	< 0,050	0,80 (0,77 - 0,83)
APACHE III	14,95	< 0,100	19,41	< 0,025	0,82 (0,80 - 0,85)
SAPS II	42,89	< 0,001	50,12	< 0,001	0,80 (0,77 - 0,83)

IC: intervalo de confianza; ROC: receiver operating curve.

TABLA 3. Razón estandarizada de mortalidad de los tres modelos

	Muertos observados	Muertos predichos	REM	IC 95%
APACHE II	262	284,61	0,92	0,83-1,01
APACHE III	262	290,66	0,90	0,82-0,99
SAPS II	262	307,40	0,85	0,77-0,93

IC: intervalo de confianza; REM: razón estandarizada de mortalidad.

fue de 3,06 días (p₂₅ 2,06 días; p₇₅ 8,28 días). La mayoría de los sujetos ingresados en el período de estudio fueron pacientes médicos (63,8%). Los resultados de la calibración y discriminación de los tres modelos se muestran en la tabla 2. Los resultados de la REM se pueden ver en la tabla 3.

Tras obtener el 10% de los pacientes por muestreo aleatorio simple, y de éstos, los pacientes en los que se podía acceder al diagnóstico de ingreso en UCI a través de la historia clínica, finalmente se incluyeron en este estudio 111 pacientes para el APACHE II y 114 para el APACHE III y SAPS II.

TABLA 4. Correlación entre los grupos 1 y 2 de intensivistas

	CCI (IC 95%)
Edad (años)	0,99 (0,98 - 0,99)
APS	
APACHE II	0,87 (0,81 - 0,91)
APACHE III	0,88 (0,88 - 0,92)
SAPS II	0,90 (0,86 - 0,93)
Puntuación de gravedad	
APACHE II	0,82 (0,75 - 0,88)
APACHE III	0,89 (0,84 - 0,92)
SAPS II	0,84 (0,77 - 0,88)
Riesgo de muerte	
APACHE II	0,75 (0,65 - 0,82)
APACHE III	0,79 (0,71 - 0,85)
SAPS II	0,80 (0,73 - 0,86)

APS: acute physiology score; CCI: coeficiente de correlación intraclase.

La correlación entre los dos grupos de intensivistas fue buena para la edad, APS, puntuación de gravedad y riesgo predicho de muerte en los tres modelos (tabla 4), (fig. 1). Para todos ellos el CCI varió entre 0,75 y 0,99.

El «t»-test para datos apareados detectó diferencias significativas entre los dos grupos de intensivistas en la puntuación de gravedad y APS, calculado para el APACHE III, así como para la puntuación de gravedad, APS y el riesgo predicho de muerte para el SAPS II. No se detectaron diferencias significativas en la edad, APS, puntuación de gravedad y riesgo predicho de muerte para el APACHE II (tabla 5).

El porcentaje de acuerdos en las comorbilidades en los tres modelos estudiados fue bueno: 86% para el APACHE II, 86,8% para el APACHE III y 94,97% para el SAPS II.

En cuanto al APACHE III el porcentaje de acuerdos en la ubicación del paciente previa al ingreso en UCI fue del 82,5%.

El porcentaje de acuerdos entre los dos grupos de intensivistas respecto del diagnóstico de ingreso en UCI para el APACHE II y el APACHE III fue bajo, del 50% en ambos casos. Sin embargo, la mayoría de los pacientes (76,58% en el modelo APACHE II y 79,82% en el modelo APACHE III) mostraron diferencias en el riesgo predicho de muerte (DR) menores del 10%. Por otro lado, la diferencia máxima de riesgo de muerte que se puede esperar como consecuencia de la diferente asignación de diagnósticos entre los dos grupos de intensivistas fue del 45% para el

DOMÍNGUEZ L ET AL. EVALUACIÓN DE LA REPRODUCIBILIDAD DE LA RECOGIDA DE DATOS PARA EL APACHE II, APACHE III ADAPTADO PARA ESPAÑA Y SAPS II EN 9 UNIDADES DE CUIDADOS INTENSIVOS EN ESPAÑA

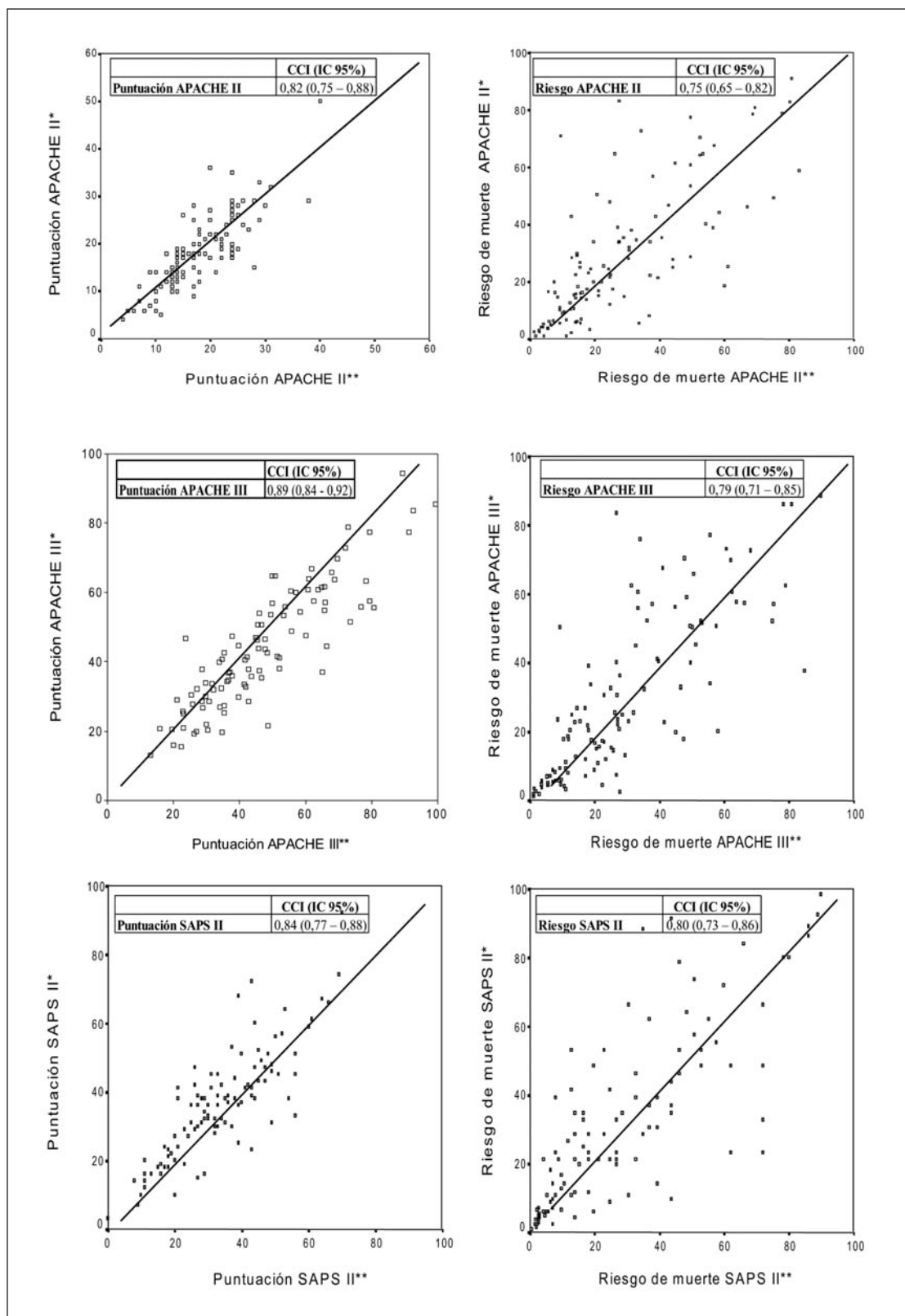


Figura 1. Correlación entre ambos grupos de intensivistas para las puntuaciones de gravedad de los modelos APACHE II, APACHE III y SAPS II (paneles de la izquierda) y para los riesgos predichos de muerte (paneles de la derecha). La línea recta de cada gráfico corresponde con la línea de identidad. *: Grupo 1 de intensivistas; **: Grupo 2 de intensivistas; CCI: coeficiente de correlación intraclase; IC: intervalo de confianza.

DOMÍNGUEZ L ET AL. EVALUACIÓN DE LA REPRODUCIBILIDAD DE LA RECOGIDA DE DATOS PARA EL APACHE II, APACHE III ADAPTADO PARA ESPAÑA Y SAPS II EN 9 UNIDADES DE CUIDADOS INTENSIVOS EN ESPAÑA

TABLA 5. Acuerdo entre los dos grupos de intensivistas

	Grupo 1*	Grupo 2*	«t» - test (IC 95%)	p
Edad (años)	62,10 (16,65)	62,04 (16,73)	0,31 (- 0,34 – 0,46)	0,760
APS				
APACHE II	11,22 (5,68)	11,44 (5,22)	-0,83 (- 0,74 – 0,30)	0,408
APACHE III	45,06 (19,42)	48,20 (20,11)	-3,48 (-4,93 - -1,35)	0,001
SAPS II	19,84 (10,71)	18,30 (10,79)	1,54 (0,65 – 2,43)	0,001
Puntuación de gravedad				
APACHE II	17,78 (7,73)	17,41 (6,74)	0,92 (-0,42 – 1,16)	0,360
APACHE III	45,21 (19,43)	48,25 (20,28)	-3,29 (-4,87 - -1,21)	0,001
SAPS II	35,55 (15,54)	32,80 (14,42)	3,44 (1,17 – 4,33)	0,001
Riesgo de muerte				
APACHE II	28,22 (22,85)	27,20 (20,22)	0,71 (-1,80 – 3,80)	0,470
APACHE III	30,01 (23,71)	29,63 (21,73)	0,27 (-2,35 – 3,12)	0,780
SAPS II	30,83 (25,16)	27,43 (23,56)	2,30 (0,55 – 6,22)	0,020

APS: acute physiology score; *media (desviación estándar).

APACHE II (en 2 de los 111 pacientes estudiados) y del 35% para el APACHE III (en 2 de los 114 pacientes estudiados). Las tablas 6 y 7 muestran el valor h con el correspondiente DR y el número de pacientes y su porcentaje en términos de frecuencia absoluta y acumulada para cada estrato de DR para el APACHE II y APACHE III.

DISCUSIÓN

En el presente estudio se ha evaluado la reproducibilidad de las puntuaciones de gravedad y sus factores principales, así como el riesgo predicho de muerte calculado por los modelos APACHE II, APACHE III adaptado para España y SAPS II. Además, se ha hecho hincapié en la evaluación de la reproducibilidad de los diagnósticos de ingreso en UCI y su impacto en la predicción del riesgo de muerte.

Tanto la puntuación de gravedad como el riesgo predicho de muerte son variables dependientes. Por lo tanto, su reproducibilidad debería depender de la de las variables independientes que se utilizan para su cálculo; estas variables son:

1. Puntuación APACHE II¹³: edad, enfermedad crónica y APS.
2. Riesgo de muerte calculado para el APACHE II¹³: puntuación APACHE II, tipo de paciente (médico o quirúrgico) y diagnóstico de ingreso en UCI.

3. Riesgo de muerte calculado por el APACHE III¹⁴: edad, comorbilidades, APS, ubicación del paciente previa al ingreso en UCI y diagnóstico de ingreso en UCI.

4. Puntuación SAPS II¹⁵: edad, APS, enfermedad crónica y tipo de paciente (médico o quirúrgico).

5. Riesgo de muerte calculado por el SAPS II¹⁵: puntuación SAPS II.

En este estudio no se encontró un desacuerdo importante entre los dos grupos de intensivistas en cuanto a la edad o enfermedad crónica en ninguno de los tres modelos. En cuanto a la ubicación del paciente previa al ingreso en UCI para el modelo APACHE III tampoco hubo desacuerdos importantes. Por lo tanto, todo apuntaba al APS y al diagnóstico de ingreso en UCI como los factores más influyentes en el desacuerdo en las puntuaciones de gravedad y riesgo predicho de muerte.

El APACHE II no mostró diferencias significativas entre los dos grupos de intensivistas en cuanto a su puntuación de gravedad, riesgo predicho de muerte y las variables independientes.

El riesgo predicho de muerte calculado por el APACHE III no fue significativamente diferente entre los dos grupos de intensivistas. Puesto que la edad y la ubicación del paciente previa al ingreso en UCI mostraron un buen acuerdo entre los dos grupos de intensivistas, quedaba por evaluar el APS y el

TABLA 6. Influencia del diagnóstico de ingreso en el riesgo predicho de muerte para el APACHE II

DR (%)	h	Pacientes absolutos	Pacientes acumulados	Porcentaje absoluto (%)	Porcentaje acumulado (%)
5	0,20	75	75	67,57	67,57
10	0,40	10	85	9,01	76,58
15	0,61	9	94	8,10	84,68
20	0,84	2	96	1,81	86,49
25	1,09	1	97	0,90	87,39
30	1,38	4	101	3,60	90,99
35	1,73	2	103	1,80	92,79
40	2,19	5	108	4,51	97,30
45	2,94	2	110	1,8	99,10

DR: diferencia de riesgo. Diferencia máxima esperada en el riesgo predicho de muerte debida a la asignación de diagnósticos diferentes a un mismo paciente por los dos grupos de intensivistas; h: diferencia máxima absoluta entre los pesos de la categoría diagnóstica asignada por los dos grupos de intensivistas. Hay un paciente con un valor h mayor de 2,94; pacientes absolutos: número de pacientes en cada intervalo; pacientes acumulados: suma de los pacientes de cada intervalo más los de los intervalos previos; porcentaje absoluto: número de pacientes en cada intervalo dividido entre el número total de pacientes; porcentaje acumulado: suma de los porcentajes de cada intervalo más los de los intervalos previos.

DOMÍNGUEZ L ET AL. EVALUACIÓN DE LA REPRODUCIBILIDAD DE LA RECOGIDA DE DATOS PARA EL APACHE II, APACHE III ADAPTADO PARA ESPAÑA Y SAPS II EN 9 UNIDADES DE CUIDADOS INTENSIVOS EN ESPAÑA

TABLA 7. Influencia del diagnóstico de ingreso en el riesgo predicho de muerte para el APACHE III

DR (%)	h	Pacientes absolutos	Pacientes acumulados	Porcentaje absoluto (%)	Porcentaje acumulado (%)
5	0,20	72	72	63,16	63,16
10	0,40	19	91	16,66	79,82
15	0,61	9	100	7,90	87,72
20	0,84	6	106	5,26	92,98
25	1,09	3	109	2,98	95,96
30	1,38	3	112	2,28	98,24
35	1,73	2	114	1,76	100,00

DR: diferencia de riesgo. Diferencia máxima esperada en el riesgo predicho de muerte debida a la asignación de diagnósticos diferentes a un mismo paciente por los dos grupos de intensivistas; h: diferencia máxima absoluta entre los pesos de la categoría diagnóstica asignada por los dos grupos de intensivistas; pacientes absolutos: número de pacientes en cada intervalo; pacientes acumulados: suma de los pacientes de cada intervalo más los de los intervalos previos; porcentaje absoluto: número de pacientes en cada intervalo dividido entre el número total de pacientes; porcentaje acumulado: suma de los porcentajes de cada intervalo más los de los intervalos previos.

diagnóstico de ingreso en UCI. El APS fue significativamente diferente. Si bien el acuerdo en el diagnóstico de ingreso en UCI fue bajo sólo en un pequeño porcentaje de pacientes, este desacuerdo afectaba al cálculo del riesgo predicho de muerte de forma importante.

En cuanto al SAPS II tanto la puntuación de gravedad como el riesgo predicho de muerte mostraron diferencias significativas entre los dos grupos de intensivistas. El acuerdo fue bueno para la edad y las enfermedades crónicas. Puesto que la otra variable independiente implicada en el cálculo del riesgo predicho de muerte es el APS, se estudió su reproducibilidad. Ésta fue mala, por lo tanto, era probable que fuera el APS la variable más influyente en la reproducibilidad del riesgo predicho de muerte por el SAPS II y su puntuación de gravedad.

Por otro lado, este estudio ha intentado buscar un método para cuantificar el impacto de la discrepancia en la asignación del diagnóstico de ingreso en UCI sobre el cálculo del riesgo predicho de muerte. Una métrica probabilística fue el método utilizado para explicar el hecho de que no todos los desacuerdos en el diagnóstico tienen el mismo impacto sobre el cálculo del riesgo predicho de muerte. En ausencia de una matriz de distancias entre los diferentes diagnósticos que se puede establecer de forma subjetiva o empírica por un grupo de expertos, nosotros desarrollamos un método para evaluar esas distancias y su impacto sobre el cálculo del riesgo predicho de muerte (anexo 1).

En este estudio el APS fue la variable más influyente en la reproducibilidad de las puntuaciones de gravedad y el riesgo predicho de muerte, como ya se ha señalado por otros autores¹⁸. Esta falta de acuerdo en el APS puede explicar la aparición de artefactos en el cálculo de la predicción de mortalidad. Errores en el cálculo, definición y exactitud de las variables, a veces por falta de experiencia, pueden influir en el cálculo de la probabilidad de muerte¹⁹. Algunos autores han demostrado cómo la implantación de un programa de entrenamiento en la recogida de los datos disminuye la variabilidad de éstos de forma considerable²⁰. También se ha apuntado que la ambigüedad en la definición de los diagnósticos puede afectar al cálculo del riesgo predicho de muerte. Por ejemplo, problemas en la definición de la ventilación mecánica

han revelado discrepancias considerables entre observadores en pacientes no intubados ventilados con presión positiva continua en la vía aérea. Otro punto importante es la valoración neurológica de pacientes sedados e intubados en los que se desconoce el estado previo de conciencia. En este caso, el modelo lo considera normal para el APS^{19,21}.

La principal limitación de este estudio deriva del hecho de que la recogida de datos se hizo de forma diferente por los dos grupos de intensivistas. El grupo 1 hizo una recogida prospectiva, y el grupo 2 una recogida retrospectiva. Esto implica problemas metodológicos a la hora de emparejar los datos. No es raro que durante el proceso diagnóstico del paciente no coincida el diagnóstico al ingreso y al alta de la UCI. Esto puede influir en que el grupo 2 de intensivistas den un diagnóstico distinto al grupo 1 en el mismo enfermo. Sin embargo, siempre se puede decir que esto es inherente y refleja fielmente el proceso clínico habitual a la cabecera del enfermo^{10,19,21}.

En conclusión, el APS y las variables que se utilizan para su cálculo fueron los factores que más influyeron en la reproducibilidad de las puntuaciones de gravedad y el cálculo del riesgo predicho de muerte para el APACHE II y APACHE III adaptado para España. La mala concordancia en el diagnóstico de ingreso en UCI no influyó de forma importante en la reproducibilidad del riesgo predicho de muerte.

Declaración de conflicto de intereses

Los autores han declarado no tener ningún conflicto de intereses.

BIBLIOGRAFÍA

- Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP. Intensive Care Society's Acute Physiology and Chronic Health Evaluation (APACHE II) study in Britain and Ireland: a prospective, multicenter, cohort study comparing two methods for predicting outcome for adult intensive care patients. *Crit Care Med*. 1994;22:1392-401.
- Apolone G, D'Amico R, Bertolini G, Iapichino G, Cattaneo A, De Salvo G, et al. The performance of SAPS II in a cohort of patients admitted in 99 Italian ICUs: results from the GiViTI. *Intensive Care Med*. 1996;22:1368-78.
- Goldhill DR, Withington PS. The effects of case mix adjustment on mortality as predicted by APACHE II. *Intensive Care Med*. 1996;22:415-9.

DOMÍNGUEZ L ET AL. EVALUACIÓN DE LA REPRODUCIBILIDAD DE LA RECOGIDA DE DATOS PARA EL APACHE II, APACHE III ADAPTADO PARA ESPAÑA Y SAPS II EN 9 UNIDADES DE CUIDADOS INTENSIVOS EN ESPAÑA

4. Cook DA. Performance of APACHE III models in an Australian ICU. *Chest*. 2000;118:1732-8.
5. Nouira S, Belghith M, Elatrous S, Jaafoura M, Ellouzi M, Boujdaria R, et al. Predictive value of severity scoring systems: comparison of four models in Tunisian adult intensive care units. *Crit Care Med*. 1998;26:852-9.
6. Markgraf R, Deuschinoff G, Pientka L, Scholten T. Comparison of Acute Physiology and Chronic Health Evaluations II and III and Simplified Acute Physiology Score II: a prospective cohort study evaluating these methods to predict outcome in a German interdisciplinary intensive care unit. *Crit Care Med*. 2000;28:26-33.
7. Moreno R. Performance of the ICU: Are we able to measure it?. En: Year book of Intensive Care and emergency medicine. Berlín, Heidelberg, New York: Springer-Verlag; 1998. p. 729-43.
8. Féry-Lemmonier E, Landais P, Loirat P, Kleinkenecht D, Brvet F. Evaluation of severity scoring systems in ICUs -translation, conversion, and definition ambiguities as a source of inter-observer variability in APACHE II, SAPS and OSF. *Intensive Care Med*. 1995;21:356-60.
9. Rowan K. The reliability of case mix measurements in intensive care. *Curr Opin Crit Care*. 1996;2:209-13.
10. Polderman KH, Girbes AR, Thijs LG, Strack van Schijndel RJ. Accuracy and reliability of APACHE II scoring in two intensive care units. *Anaesthesia*. 2001;56:47-50.
11. Rué M, Valero C, Quintana S, Artigas A, Álvarez M. Inter-observer variability of the measurement of the mortality probability models (MPM II) in the assessment of severity of illness. *Intensive Care Med*. 2000;26:286-91.
12. Polderman KH, Christians HM, Wester JP, Spijkstra JJ, Girbes AR. Intra-observer variability in APACHE II scoring. *Intensive Care Med*. 2001;27:1550-2.
13. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;13:818-29.
14. Rivera-Fernández R, Vázquez-Mata G, Bravo M, Aguayo-Hoyos E, Zimmerman J, Wagner D, et al. The APACHE III prognostic system: customized mortality predictions for Spanish ICU patients. *Intensive Care Med*. 1998;24:574-81.
15. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA*. 1994;270:2957-63.
16. Shorack GR, Wellner JA. Empirical processes with applications to statistics. New York: Wiley and Sons; 1986.
17. Munk A, Czado C. Nonparametric validation of similar distributions and assessment of goodness of fit. *J R Stat Soc B*. 1998;60:223-41.
18. Polderman KH, Thijs LG, Girbes AR. Interobserver variability in the use of APACHE II scores. *Lancet*. 1999;353:380.
19. Liddy M, Chen MB, Martin CM, Morrison TL, Sibbald WJ. Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. *Crit Care Med*. 1999;27:1999-2004.
20. Kees H, Polderman E, Jorna MF, Girbes ARJ. Inter-observer variability in APACHE II scoring: effect of strict guidelines and training. *Intensive Care Med*. 2001;27:1365-9.
21. Schuster DP. Predicting outcome after ICU admission. The art and science of assessing risk. *Chest*. 1992;102:1861-70.