

Validación del APACHE II y SAPS II en la República Argentina

AUTORES: *Comité de Scores de S.A.T.I. e Investigadores.*

(Comité de Scores: Cueto, Graciela; Torres Boden, Margarita; Vetere, Liliana; Santos, Cristina; Castarataro, Carlos; Pimentel, Rodrigo; Wolanow, Viviana.)

RESUMEN

INTRODUCCION: Los score s fueron diseñado s para poder predecir la mortalidad de pacientes críticos, considerando las características de los pacientes que podrían influir en su riesgo de muerte. Para poder aplicar dicho s score s en un a población di s tinta de la base original debe establecerse si es tan confiable como en la población original. Una de las formas de constatarlo es por medio de la validación.

OBJETIVO: comparar e l rendimiento de l APACHE II (Acute Physiology and Chronic Health Evaluation) y SAPS II (A New Simplified Acute Physiology Score) en la República Argentina y constatar si uno de ellos es superior en nuestra población.

MATERIAL Y MÉTODOS:

Diseño: cohorte prospectiva y multicéntrica.

Lugar: 36 unidades de terapia intensiva de 14 provincias de la República Argentina.

Pacientes: ingresaron 7.127 paciente s admitido s consecutivamente en un período de 15 meses.

Criterios de inclusión: paciente s de 16 años de edad o mayores, ingresados en forma consecutiva y con un a permanencia igual o mayor a 24 horas en la Unidad de Terapia Intensiva.

Métodos: se recogió la información necesaria para computar lo s score s APACHE II y SAPS II como lo realizaron lo s autores originale s y se aplicaron sus fórmulas. El pronóstico evaluado fue la mortalidad hospitalaria. Se establecieron cuatro estratos de severidad de acuerdo al valor de APACHE II. Se utilizó para las mediciones estadísticas el programa SPSS 10.1 for Windows, incluyendo la calibración (goodness of fit) y la discriminación (area under the receiver operating characteristic curve).

RESULTADOS: la edad media fue de 54.8 ± 19.5 . El motivo de ingreso fue clínico en el 59.8%, y cirugía de urgencia en el 20.9%. La mortalidad global fue del 28%. La predicha por el APACHE II fue 24.5% y por el SAPS II de 19.2%. Ambos score s subestimaron el riesgo de muerte en la Argentina. Tanto

ABSTRACT

INTRODUCTION. Scoring systems were designed to predict the evolution of groups of critical care patients, taking into account variables that could affect the risk of death. However, they are referred to a certain place and time. To support our use of them, we need to establish if the predictive performance in the new population is so confident as in the original one. This can be done with a validation process.

OBJECTIVES. To compare the performance of the two severity-of-illness scores APACHE II and SAPS II used in ICU patients in Argentina, and to establish if one of them might be superior in our population.

MATERIALS AND METHODS. Design: multi-centre, national, prospective cohort study. Setting: 36 ICUs in 14 provinces in Argentina. Patients: All consecutively admitted during a 15-month period, > 16 years old and with an ICU stay > 24 hs. Method: Collection of information required to compute APACHE II and SAPS II scores, as in the original studies. The outcome measure was hospital mortality. Four severity-grading values of APACHE II were settled. Calibration (goodness of fit) and discrimination (area under the receiver operating characteristic curve) were calculated.

RESULTS. 7.128 patients from 36 institutions were included. Mean age was 54.8 ± 19.5 . 59.8% of patients were non-surgical, and 20.9% had emergency surgery. Global was 28%. Predicted mortality was 24.56% with APACHE II and 19.28% with SAPS II. Both scores underestimated the hospital mortality risk in our population. APACHE II and SAPS II showed a poor calibration. Discrimination showed a better performance, with receiver operating characteristic areas above 0.795. Comparison between both curves was similar. For APACHE II score, sensitivity was 40.79%, specificity was 91.87% and rank of correct classification was 77.53%. For SAPS II score, sensitivity was 42.21%, specificity 93.67%

el APACHE II y SAPS II mostraron una mala calibración. La discriminación fue superior con áreas bajo la curva de ROC superiores. La comparación entre ambos scores fue similar. En la tabla de 2x2 la sensibilidad fue del 40.79%, la especificidad del 91.87% y el rango de clasificación correcta de 77.53% para el APACHE II. Para el SAPS II la sensibilidad fue 21%, la especificidad de 93.67% y clasificó correctamente al 79.26% de los pacientes.

CONCLUSIONES: el APACHE II y SAPS II no validan adecuadamente en la Argentina. La calibración para ambos scores es pobre. La discriminación es superior logrando una mayor separación entre los pacientes vivos y muertos de acuerdo a los scores. No hay diferencias significativas entre los dos scores. Ambos subestiman la mortalidad en la población argentina. Sin embargo, el APACHE II estratifica adecuadamente a los pacientes de acuerdo a su gravedad y permite que sea usado para investigación clínica.

and rank of correct classification was 79.26%. In subgroup analysis, surgical population showed better calibration than the whole population, and discrimination was still good.

CONCLUSIONS. Both predictive models showed similar performance. They demonstrated poor calibration for the whole database. Discrimination was slightly better with SAPS II, but there was no statistical difference between both scores. Hospital mortality was higher than predicted with both scores, but underestimation was largest in the middle-range risks. The stratification of severity provided by APACHE II is good and seems to allow its use for clinical investigation in Argentina.

INTRODUCCION

En las últimas décadas, se ha invertido tiempo, investigación e importantes recursos económicos en la búsqueda de herramientas que permitan predecir certeramente la evolución de los pacientes críticos y brindar un cuidado más racional y costo efectivo⁽¹⁻²⁻³⁻⁴⁾

La edad crecientemente de la población, el aumento del número de camas en las Unidades de Terapia Intensiva (UTI), el incremento de los costos y la limitación de los fondos, lleva a una mayor necesidad de establecer parámetros objetivos que evalúen la severidad de la enfermedad.⁽⁵⁻⁶⁻⁷⁾

Se han desarrollado sistemas para analizar la severidad de la enfermedad y medir la mortalidad hospitalaria de los pacientes críticos.⁽³⁻⁸⁻⁹⁻¹⁰⁻¹¹⁻¹²⁻¹³⁾

Dentro de estos sofisticados modelos predictivos, se encuentran los scores inespecíficos que evalúan a la población general de terapia intensiva, y fisiológicos que miden la respuesta común a diferentes injurias.

En 1981, fue introducido el score APACHE⁽⁸⁾ (*Acute Physiology and Chronic Health Evaluation*) desarrollado por Knauer y cols. Se usó un método subjetivo para la selección de las variables y para determinar su puntaje. Su realización dificultó a para la práctica clínica impidió su generalización. En 1984, Le Gall y cols⁽⁹⁾ introdujeron un score simplificado del APACHE original, que se denomina SAPS (*Simplified Acute Physiology Score*). Consta de 13 variables fisiológicas y la edad. Al año siguiente, Knauer y cols.⁽¹⁰⁾, publicaron una reducción del sistema original, denominada APACHE II. Abrevian las variables fisiológicas a 12, combinando la evaluación del estado de salud

previo del paciente y la edad, con el grado de deterioro fisiológico en las primeras 24 horas de internación en UTI. Calcula la probabilidad de sobrevida individual mediante un modelo de regresión logística, agregando el peso específico considerado para un diagnóstico. De la sumatoria de todos ellos se obtiene la predicción para esa población de 5815 pacientes. También el grupo europeo, en 1993, da a conocer el SAPS II, desarrollado en un grupo de 8549 pacientes y validado internamente en otros 4.603⁽¹³⁾. Utiliza la regresión logística múltiple en la selección de variables, sus rangos y los puntos asignados. En la estimación de la probabilidad de mortalidad no considera el diagnóstico de ingreso. Estos dos últimos, el APACHE II y el SAPS II son los más utilizados para la predicción en UTI.

Estos modelos predictivos invariablemente reflejan las características de la población original y la cultura médica del lugar, por lo que su transferencia a otros centros se encuentra viciada.⁽¹⁰⁻¹⁴⁻¹⁵⁻¹⁶⁻¹⁷⁾ Sin embargo, es necesaria la comparación con modelos universales.

Al comparar, es necesario verificar si los valores establecidos en la población patrón son válidos en un grupo diferente de pacientes⁽⁷⁾. Este proceso se conoce como validación externa. Utiliza dos técnicas estadísticas. Una de ellas, la calibración, constata si la mortalidad predicha por el modelo se corresponde con la mortalidad real observada⁽¹⁸⁾. Otra técnica es la discriminación, que valora si el modelo propuesto tiene la capacidad de diferenciar los vivos y los muertos en la nueva muestra⁽¹⁹⁾.

En este estudio utilizamos ambas técnicas para realizar la validación del APACHE II y SAPS II

TABLA I: Variables globales

Variable	N/%
Sexo masculino	4.131 (57,9)
Cirugía electiva	1.371 (19,2)
Cirugía de urgencia	1.489 (20,9)
Clínicos	4.259 (59,8)
Mortalidad en UTI	1.734 (24,3)
Mortalidad hospitalaria	1.995 (28,0)
Readmisiones	364 (5,0)
Edad	54,8 ± 19,51
Estadía en UTI	6,96 ± 10,63
APACHE II	15,12 ± 8,84
SAPS II	31,73 ± 17,6

Expresadas como media y desvío standard (\pm SD)

en la República Argentina (RA). Nuestros objetivos fueron:

- 1) Validar los scores inespecíficos APACHE II y SAPS II en la República Argentina (RA).
- 2) Comprobar si alguno de ellos es superior en nuestra población.
- 3) Describir las características demográficas de la población de pacientes ingresados a las UTIs de la RA.

MATERIAL Y METODOS

Diseño: estudio de cohorte y prospectivo, multicéntrico y nacional.

Criterios de inclusión:

- 1) Pacientes de edad mayor o igual a 16 años ingresados en forma consecutiva.
- 2) Permanencia mayor o igual a 24 horas en la Unidad de Terapia Intensiva.

Criterios de exclusión:

- 1) Pacientes menores a 16 años.
- 2) Permanencia menor a 24 horas en la Unidad de Terapia Intensiva.
- 3) Quemados.

Metodología:

Se remitió la invitación para participar a 194 centros de la RA. Para la selección de centros se tomó en cuenta que los mismos tuvieran nivel de complejidad VII I (cuenta con diversas especialidades y servicios auxiliares de complejidad) y IX (se suma a lo anterior que constituyan centros de docencia universitaria, en ambos casos de derivación regional) y que estuvieran adscriptos a la Sociedad Argentina de Terapia Intensiva (SATI). Respondió el 27,08 % (52 UTI). Finalmente 36 centros participaron de la muestra, durante un período de 15 meses consecutivos, desde el 1 de marzo de 1999 hasta el 31 de mayo de 2000.

La recolección de datos se hizo mediante planillas estandarizadas, con instructivo adjunto y estableciendo criterios para aquellos parámetros que pudieran presentar controversia y que no fueron aclarados por el autor en el APACHE II (Apéndice 1)

En el Centro Coordinador se incluyeron los valores en un software de base de datos unificada. El programa realizado para el estudio utilizó, en las fórmulas de regresión logística, los coeficientes originales, tanto del APACHE II como del SAPS II, para la predicción de mortalidad. Se estableció el tratamiento de los datos perdidos, excluyendo los mismos, en la metodología de análisis. Durante todo el estudio, los centros participantes, además de contar con el instructivo podían contactarse con los coordinadores. Al finalizar la recolección central, se reenviaron para su corrección aquellas planillas con datos relevantes ausentes. Se evaluaron tendencias estadísticas a los 1200 pacientes, revisando la metodología informática.

Se consideraron las variables consignadas en los estudios originales de Knaus y col⁽¹⁰⁾ y de LeGall y col⁽¹³⁾, utilizando las instrucciones y el formato establecido por ellos. Con respecto al diagnóstico de ingreso se utilizaron las 50 categorías establecidas por Knaus y col⁽¹⁰⁾, estudiándose su prevalencia. Se consignaron los datos demográficos que permiten la descripción de la población argentina. Se establecieron 4 estratos de severidad, de acuerdo al incremento de la tasa de mortalidad y su diferencia estadística, para estimar la escala de gravedad de los pacientes internados. Estos rangos fueron establecidos con valores de APACHE II de 0-9 puntos (p), 10-19 p, 20-29 p y mayor de 30 p. Fue registrada la mortalidad de UTI y hospitalaria.

Método Estadístico: la variable dependiente fue la mortalidad hospitalaria. El análisis estadístico se efectuó por medio de regresión logística múltiple, en el programa SPSS 10.1 para Windows (SPSS Inc. Chicago, IL 2001). Las variables categóricas se expresan como frecuencia y porcentaje, mientras que las numéricas como media y desvío standard (SD). El valor de significación estadística fue establecido para una $p < 0,05$ para dos colas.

Uso de Regresión Logística (RL) para armar el modelo predictivo en los scores APACHE II y SAPS II: la Regresión Logística Múltiple, es un método estadístico que permite obtener una predicción de que un evento ocurra. Es una fórmula matemática que incorpora las características del paciente que influyen en el evento estudiado (variable dependiente). En este caso, las variables fisiológicas sumadas en un score (variables independientes), más la condición de ingreso a la UTI (clínico o quirúrgico) y en el caso del APACHE II, el principal diagnóstico de ingreso, no incluido en el SAPS II. Para asignarle el peso o influencia que tienen cada uno de ellos en la mortalidad hospitalaria, que es el

evento estudiado en estos dos modelos, cada una de estas variables independientes se multiplican por un coeficiente derivado de la base de datos original del modelo. El valor obtenido por esta fórmula es la probabilidad de mortalidad hospitalaria (probabilidad de que ese evento ocurra) (Apéndice 2).

Esta probabilidad de morir o no sobrevivir (NS), durante la estancia hospitalaria se expresa por un valor que oscila de 0 a 1. Cuando ese valor se acerca a 0 significa que tiene una alta probabilidad de sobrevivir y si el valor se acerca a 1, el paciente tiene una alta probabilidad de morir. Si consideramos un punto de corte de 0,5 significa que el paciente que tiene una probabilidad de mortalidad hospitalaria > a 0,5 y se lo clasifica como no sobreviviente; y si es < a 0,5 lo clasifica como sobreviviente. La probabilidad de sobrevivir se obtiene para cada paciente, para luego aplicar a la población se debe hacer la sumatoria de las probabilidades individuales. La probabilidad de mortalidad se calculó de acuerdo a las ecuaciones de los trabajos originales.

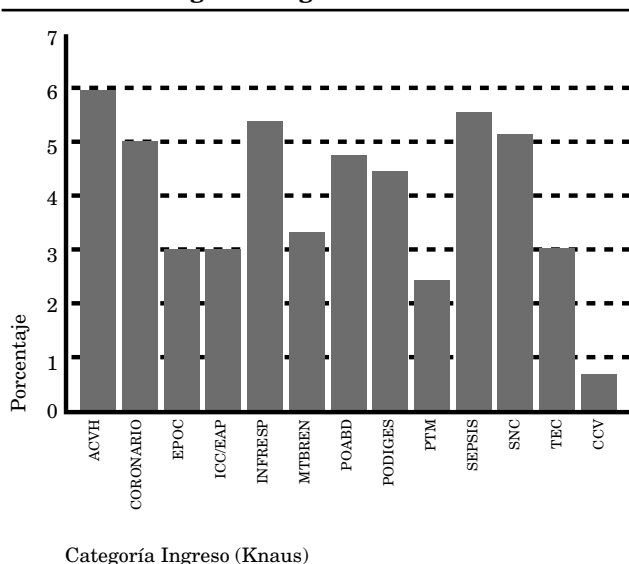
La calibración de ambos se estudió por medio del método estadístico Goodness-of-Fit (GOF) de Hosmer y Lemeshow. Este método permite comprobar si los resultados pronosticados por el modelo (sobrevivida -SV- o no sobrevivida -NS) para cada paciente individual se corresponden con el resultado real observado. Una buena calibración, significa que ambos resultados no tienen diferencias significativas (lo pronosticado versus lo observado o real) y está representado por un GOF inferior a 15.5.⁽⁷⁻²⁰⁾ Otra forma de decirlo es que, el modelo pronosticó un resultado que se acerca a lo que realmente sucede en esa población. Se establecen distintos niveles de riesgo de mortalidad y en cada uno de ellos se compara lo pronosticado con lo observado obteniéndose un chi-cuadrado para cada uno de ellos.

La discriminación investigada la habilidad de un modelo para distinguir entre un paciente que sobrevivirá y uno que fallecerá. Estudia la sensibilidad y especificidad del score en diferentes puntos de corte de probabilidad de mortalidad construyendo la curva Receiver Operating Characteristic (ROC) y calculando el área bajo la curva (AUC). Cuando la predicción es casi perfecta el número se acerca a uno, y cuando es producto del azar se aproxima a 0.5. Se considera que una buena discriminación debe ser superior a 0.80. Se calcularon la Sensibilidad (S), la Especificidad (E) y el rango de clasificación correcta (RCC) a un corte de 0.5 de probabilidad de supervivencia, clasificándolos según la tabla de 2x2 de outcome.⁽¹⁹⁾

RESULTADOS

Se incluyeron 7.128 pacientes ingresados consecutivamente a 3 centros, correspondientes a 11 provincias (Apéndice 1). Las características gene-

Figura 1. Frecuencia de motivos de ingreso según categorías de Knaus



rales de la población se observan en la TABLA I. El 57.9% correspondió al sexo masculino, el 59.8% ingresó por su condición clínica. De los pacientes quirúrgicos, 1.489 (20.9%) correspondieron a cirugía de urgencia. El 5% de pacientes reingresaron. La edad media fue de 54.8 ± 19.5 y la estadía promedio en UTI de 6.96 ± 10.63 días. La estadía hospitalaria no se pudo obtener por omisión en la recolección del dato en el 15.2% (n = 1083) de los formularios.

Los motivos de ingreso más frecuentes según las categorías diagnósticas de Knaus fueron: accidente cerebrovascular hemorrágico 6.0%, sepsis 5.6%, insuficiencia respiratoria 5.4%, coronariopatía 5.0% y patología clínica del sistema nervioso central 5.0%.

La mortalidad global fue del 28.0%, la predicha por el APACHE II del 24.6% y por el SAPS II 19.3% (TABLA II). Ambos scores subestimaron el riesgo de mortalidad hospitalaria en la RA, aunque lo fue en mayor medida por el SAPS II.

Las medias de la edad, estadía y los valores de APACHE II y SAPS II fueron significativamente superiores para los no sobrevivientes (TABLA III).

TABLA II: Resultados relacionados con la mortalidad

Variable	n	%
Mortalidad Real	1.995	28.0
Mortalidad Predicha por APACHE II	1.743	24.6
Mortalidad Predicha por SAPS II	1.423	19.3

TABLA III: Diferencia entre sobrevivientes y fallecidos

Variable	SOBREVIVIENTES		NO SOBREVIVIENTES		Significación P < 0.05*
	Media	± SD	Media	± SD	
Edad	52,63	19,61	60,55	18,05	0.001
Estadía UTI	6,25	8,87	8,77	14,02	0.001
APACHE II	12,54	7,05	21,73	9,51	0.001
SAPS II	26,36	14,54	45,52	17,36	0.001

* Test U de Mann-Whitney

TABLA IV: Grupos de APACHE II

Grupos de A II	n	Frecuencia %	Promedio de A II *	Mortalidad UTI % (n)	Mortalidad Hospitalaria % (n)	Significación p < 0.05**
I: 0-9	2084	29.2	6	3.7 (77)	5.6 (117)	0.001
II: 10-19	3051	42.8	14	18.8 (549)	23.6 (720)	0.001
III: 20-29	1555	21.8	23.6	49.3 (766)	53.4 (830)	0.001
IV: >30	438	6.1	34.5	72.4 (317)	75.1 (329)	0.001

*AII: APACHE II;

**Test de Kruskal-Wallis

Según división por gravedad el 29.2 % pertenece al grupo I, el 42.8 % al grupo II, el 21.0 % al grupo III y el 6.1 % al grupo IV. La mortalidad hospitalaria fue del 5.6%; 23.6%; 53.4% y 75.1% respectivamente. Estos resultados muestran que efectivamente el incremento de puntaje de APACHE II se correlaciona con aumento en la mortalidad ($p < 0.001$) (TABLA IV).

La calibración del APACHE II dio un χ^2 para el GOF de 92.89. En la tabla de clasificación de pronósticos observados y esperados (TABLA V), encontramos una sobrestimación de la mortalidad en los 3 primeros decilos, así como en los dos últimos. En los estratos intermedios se ve subestimada. La discriminación evidenció un área bajo la curva ROC de 0.795 con intervalo de confianza de 0.784 a 0.807 (Figura 3).

El SAPS II mostró un χ^2 de 151,17. Al analizar los estratos de gravedad nuevamente se observó una sobrestimación en los cuatro primeros niveles, de allí pasa a subestimar hasta el último decilo (TABLA VI). En este caso, la discriminación mostró un área bajo la curva ROC de 0.814 con un intervalo de confianza que va del 0.803 al 0.825. (Figura 4). La comparación entre las dos curvas ROC no muestra diferencia significativa.

En la tabla de 2x2 se encuentra que para el criterio de decisión del 0.5 en el APACHE II la Sensibilidad es del 91.87% y la tasa de clasificación correcta es del 77.53%. Para el SAPS II la Sensibilidad es del 42.21%, la Especificidad del 93.67% y un rango de clasificación correcta de 79.26% (TABLA VII).

En el análisis de subgrupos, de acuerdo a la categoría diagnóstica, encontramos que los pacientes quirúrgicos tienen mejor calibración que la población general, sobre todo con el APACHE II, con una discriminación que continuó siendo buena, logrando un RCC superior al 80%. El mismo comportamiento se repite en otras patologías como el accidente cerebrovascular hemorrágico. El grupo que logra la mejor predicción es post-operatorio de abdomen con un GOF (AII 10.96; SII 11.75) y área bajo la curva ROC muy buenos por ambos sistemas. El SAPS II presenta un comportamiento paralelo en los subgrupos de tipo de ingreso y patologías (TABLA VIII).

DISCUSION

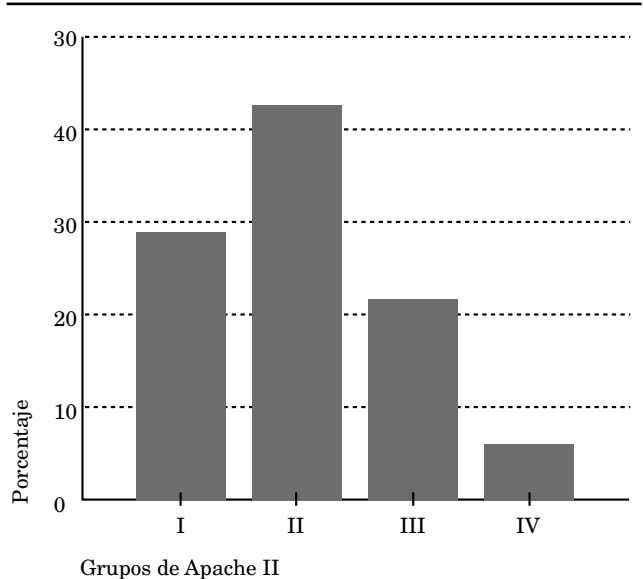
La posibilidad de predecir el pronóstico o riesgo de morir para un grupo de pacientes crítico es un elemento indispensable en el manejo de las UTIs.⁽⁵⁾ Se han utilizado distintos scores con este fin, pero

dentro de ellos se destacan algunos. El APACHE II, pese a ser de los primeros y haber sido superado en diseño por las nuevas generaciones de scores, continúa siendo el de uso más generalizado como herramienta en la valoración de la gravedad. En el continente europeo, el SAPS II, diseñado en años posteriores, con un método más sofisticado, ha ocupado un lugar similar.

El objetivo propuesto en nuestro trabajo, fue comprobar si era posible el uso de ambos o alguno de ellos para establecer pronósticos en la población argentina mediante el proceso de validación externa.⁽⁷⁻²⁰⁾

Los esfuerzos previos realizados en este sentido demostraron que cada score depende de la base de desarrollo para obtener su mejor predicción.⁽²¹⁾ Zimmerman y col⁽¹⁴⁾, uno de los autores del APACHE II, señala que su mayor limitación al ser trasladado a otras poblaciones, es establecer una predicción de mortalidad con datos obtenidos en Estados Unidos durante los años 1979 a 1982. Se trata de explicar esta diferencia en distintas formas. Una de ellas, es atribuida a la diferente mezcla de casos diagnósticos propia de cada unidad.⁽¹⁴⁻²²⁻²³⁾ Este es, si embargo, uno de los fundamentos del uso de los scores inespecíficos, ya que los pacientes críticos comparten el deterioro fisiológico común que las distintas enfermedades producen.⁽²⁴⁾ Otra causa sugerida es que las características y la práctica clínica de cada UTI no es tomada en cuenta por estos scores (proporción de pacientes clínicos, criterios de admisión, de derivación, etc.).⁽¹⁷⁾ Asimismo, la falta de ajuste se puede presentar cuando predominan ciertas enfermedades que no son adecuadamente valoradas y están representadas significativamente en la base original, como el trauma⁽²⁵⁻²⁶⁻²⁷⁾ y la cirugía cardiovascular⁽²⁸⁻²⁹⁾. Livingston e⁽³⁰⁾ estudió la performance de cinco modelos predictivos en 10.393

Figura 2. Grupos de Apache II. Frecuencia.



pacientes de 22 Unidades de Escocia. Los resultados fueron uniformes entre todos ellos. Logra una discriminación razonable, pero la calibración es inadecuada sin lograr la validación de ninguno de los scores utilizados. Este es el resultado que se obtiene con este tipo de estudios: la discriminación es superior a la calibración.⁽¹⁰⁻¹⁴⁻²²⁾ Cuando la nueva población forma parte de las bases de desarrollo, esta diferencia se reduce. Castella y col⁽¹⁵⁾ al compararon estos mismos scores y encontraron diferencias entre el APACHE II y el SAPS II a favor de este último. Si embargo, más del 70% de los pacientes en su base de datos pertenecían, a los centros de desarrollo del SAPS, explicando su superioridad.

TABLA V: Test de Goodness-of-Fit de Hosmer Lemeshow del APACHE II

Grupos	Sobrevivientes		No Sobrevivientes		Total
	Observados	Esperados	Observados	Esperados	
1	685	639.65	27	72.36	712
2	657	633.40	56	79.60	713
3	649	625.17	64	87.83	713
4	602	611.12	109	99.88	711
5	583	592.10	126	116.90	709
6	531	566.32	179	143.68	710
7	495	525.08	215	184.92	710
8	399	448.07	311	261.93	710
9	318	315.69	391	393.32	709
10	186	148.40	515	552.59	701
$\chi^2 = 92.89$		df = 8	Sig= .0000		

Figura 3. ROC Apache II

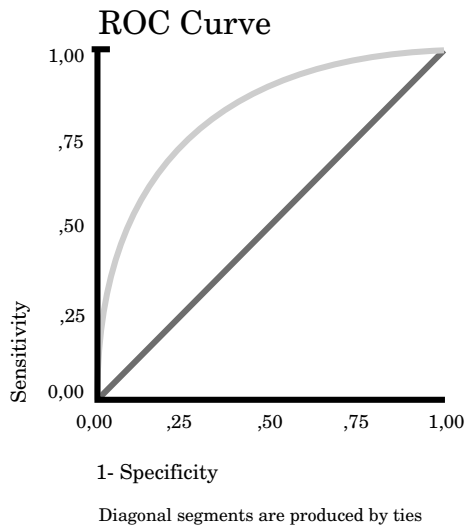
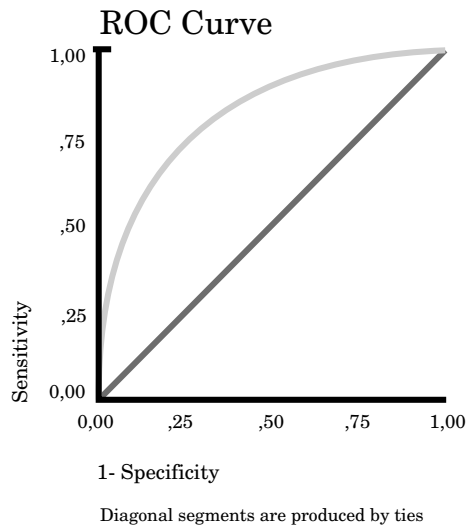


Figura 4. ROC SAPS II



En nuestro estudio se repite este comportamiento ya que ni el APACHE II, ni el SAPS II, valían en la población de la RA. Al evaluar la calibración, observamos que ambos scores subestiman del riesgo de morir de manera global, y más precisamente en los estratos intermedios. En los extremos de gravedad, ambos scores sobreestiman la mortalidad. La discriminación, valorada por el área bajo la curva ROC, tiene un valor entre 0.79 y 0.81, es decir que la discriminación entre vivos y muertos es superior a la calibración, pudiendo calificarse su valor como bueno. Nuevamente constatamos que los modelos exportados no logran brindar una cierta predicción.⁽³¹⁾

Se puede especular que las causas de este resultado se deben a la diferente mezcla de casos general (la proporción de pacientes clínicos y quirúrgicos de urgencia que difieren con respecto a la base original); la diferente mezcla de casos en cada estrato de gravedad; las diferentes proporciones de categorías diagnósticas de pacientes y sin duda, diferencias en las prácticas como los criterios de atención, internación y egresos. A priori, se tiende a pensar que esta diferencia de resultados significa una atención de los pacientes críticos en la Argentina, sin embargo, debe ajustarse previamente por esas diferencias para calificar un resultado con respecto al otro.

Al comparar ambos scores entre sí, para seleccionar uno de ellos para nuestro país, no encontramos diferencias significativas entre ambos, si bien el SAPS II logra mejor discriminación. Este último, pese a tratarse de un score muy superior al APACHE II en su diseño estadístico y su validación, no logra superarlo significativamente. Por lo tanto, no es posible dar una recomendación en cuanto a la

selección de uno de ellos. Ambos pueden ser usados conociendo su valor limitado en nuestro país.

La estratificación en grupos con distintos niveles de gravedad de mejor a el enfoque objetivo, sirve cuando queremos aplicar el método científico en estudios randomizados, al estudiar una terapéutica, al describir una evolución y al estimar la mortalidad. Identificar estos grupos de riesgo ingresados a cada Unidad, es útil en la distribución de los recursos requeridos para su atención, y a sean humanos, económicos, edilicios, administrativos, etc. El APACHE II es eficiente en clasificar de esta manera a los pacientes⁽¹⁵⁾ aún cuando la validación no sea óptima. Nuestros resultados muestran un incremento significativo de la mortalidad al aumentar el puntaje. También otros, como Angus y col⁽³⁰⁾ observaron, en un estudio sobre trasplantes hepáticos, que a pesar de una inadecuada discriminación, estratificaba adecuadamente a los pacientes por gravedad.

Se encuentra un alta proporción de pacientes con puntajes bajos, probablemente debido a la ausencia de unidades intermedias y la necesidad de monitorear pacientes de riesgo potencial, en especial, postoperatorios inmediatos. Probablemente, si analizáramos por tipo de institución, públicas o privadas, encontraríamos que en las segundas hay predominancia de pacientes de baja gravedad que no requerirían la complejidad y el costo de UTI.

Para superar esta dificultad que se presenta al trasladar los scores, se propone la personalización o adecuación de los modelos predictivos a la región estudiada, es la llamada "customización", aún no realizada con esta base. Es posible hacerla de distintas maneras. Una de ellas, es el cambio de los

coeficientes originales por otros obtenidos de la nueva base. Livingstone, dentro de los 5 scores que utilizó, también seleccionó el APACHE II "customizado" en Gran Bretaña⁽³⁰⁾. Sirió⁽³²⁾ en Cleveland, lo personalizó para su estado. Sin embargo, en ambas experiencias tampoco lograron una adecuada validación. Esto implica que es necesario continuar con estudios que permitan mejorar la capacidad predictiva en los pacientes críticos.

¿Cuál sería entonces la importancia de insistir en el uso de scores que no logran su cometido? Lo relevante es que son parámetros objetivos, a pesar de sus limitaciones y para utilizarlos como una herramienta válida, habría que modificar su uso.

Cuando los aplicamos en la RA pueden ser usados para comparar el rendimiento en nuestras Unidades, ya sea en un mismo centro a través del tiempo, o entre distintas instituciones en un mismo período. Se deberán señalar las diferencias ajustadas de sus valores cuando las apliquemos en comparaciones en el exterior. Deberán también complementarse con otras herramientas según el objetivo propuesto, considerando si evaluamos calidad, costos o satisfacción.⁽¹⁷⁻²⁹⁾

CONCLUSIÓN:

Este estudio representa la población argentina, ya que participaron 11 provincias, importante cantidad de centros y alcanzó un número relevante de pacientes. La mezcla de casos es real y no se ve influida por el comportamiento de un grupo en particular. Al haberse realizado en un período de 15 meses, no se ve sesgado por la variación de la conducta médica.

Aporta la descripción de las variables fisiológicas, del tipo de pacientes ingresados a las Unidades de Terapia Intensiva en la Argentina, su motivo de ingreso, mortalidad y estadía en UTI, convirtiéndolo en una invaluable herramienta que permite establecer comparaciones en nuestro país.

Nuevamente, se constata que la comparación con modelos universales no es posible si no haber sido antes sometido a la validación externa. Los pacientes de la RA no pueden ser analizados en sus pronósticos siguiendo los criterios establecidos por el APACHE II y el SAPS II. Sin embargo, en tanto son scores inespecíficos, pueden ser utilizados para estratificar la gravedad de los pacientes para estudios clínicos, auditorías, etc. Y definitivamente pueden usarse para establecer comparaciones fide-

TABLA VI: Test Goodness-of-Fit de Hosmer y Lemeshow del SAPS II

Grupos	Sobrevivientes		No Sobrevivientes		Total
	Observados	Esperados	Observados	Esperados	
1	690	643.04	34	80.96	724
2	694	654.14	48	87.86	742
3	657	625.28	59	90.72	716
4	679	668.54	97	107.46	776
5	589	604.50	128	112.50	717
6	580	628.89	196	147.11	776
7	480	544.79	251	186.21	731
8	401	426.11	296	270.89	697
9	246	262.66	454	437.35	700
10	115	73.06	432	473.94	547
$\chi^2 = 151,17$		df = 8		Sig= .0000	

Tabla VII: Performance de los modelos

	Mortalidad Predicha %	GOF	ROC	IC 95%	Sensibilidad	Especificidad	RCC %
APACHE II	24.56	92.89	0.795	0.784-0.807	40.79	91.87	77.53
SAPS II	19.98	151.17	0.814	0.803-0.825	42.21	93.67	79.26

TABLA VIII : Análisis de subgrupos del APACHE II Y SAPS II: "Calibración y Discriminación"

Subgrupos	n	APACHE			S	E	RCC
		II	GOF	ROC			
Clínicos	3.907	16.5	94.14	0.79	44.6	90.4	75.8
Quirúrgicos	2.614	13.5	19.64	0.78	36	94	81.3
A.C.V.H. ¹	379	17	8.37	0.81	72.3	77.6	75
P CL SNC ²	833	16.9	26.4	0.80	48.9	77	64.2

¹Accidente Cerebro Vascular Hemorrágico; ²Patología Clínica del Sistema Nervioso Central

Subgrupos	n	SAPS II			S	E	RCC
		II	GOF	ROC			
Clínicos	3.907	34.2	103.2	0.80	46.5	91.6	77.2
Quirúrgicos	2.614	28.2	45.05	0.81	39.4	95.3	82.9
A.C.V.H. ¹	379	36.5	16.7	0.82	64.4	83.9	74.6
P CL SNC ²	833	36.7	13.97	0.79	51.3	84.7	70.8

dignas dentro de la RA. Éste ha sido el primer paso, falta aún, completar el camino para lograr un ajuste más adecuado superando las limitaciones evidenciadas por este análisis.

APÉNDICES

Apéndice 1: Aclaraciones en la confección de planillas

1) Frecuencia respiratoria:

- ❖ Si el paciente se encuentra en Ventilación Mecánica durante las primeras 24 horas de internación en Terapia Intensiva, se considerará para la realización del puntaje la causa de ingreso a ARM. Si fuera por causa anestésica o neurológica, considerar 0 punto, y si fuera por causa cardiorrespiratoria considerar 4 puntos.
- ❖ Si antes de las 24 horas de ingreso se retira el ventilador, se debe considerar el puntaje de la frecuencia respiratoria espontánea posterior.
- ❖ Si antes de colocar el ventilador existió un período de ventilación espontánea, considerar la frecuencia de dicho período para tomar el puntaje

2) Glasgow:

- ❖ Respecto de los pacientes que requieren sedación y/o relajación, considerar el Glasgow previo a la sedación.
- ❖ En pacientes en postoperatorio inmediato que requieran mantener la sedación por el tratamiento de su enfermedad de base, y en los que no se suspende la sedación por más de 24 horas, considerar el Glasgow previo a la cirugía. Esto es válido tanto para las cirugías neurológicas, como para extra-neurológicas.

Apéndice 2: Cálculo de la Probabilidad de Mortalidad Hospitalaria

1) Probabilidad de Mortalidad Hospitalaria

$$= \frac{e^{\text{logit}}}{1 + e^{\text{logit}}}$$

Logit = -3.517 + (APACHE II x 0,146) + (0,603 si el ingreso es por cirugía de urgencia) + (coeficiente beta del motivo de ingreso a UTI)

e = base de los logaritmos naturales (2,718)

2) Probabilidad de Mortalidad Hospitalaria

$$= \frac{e^{\text{logit}}}{1 + e^{\text{logit}}}$$

Logit = -7,7631 + 0,0737 x SAPS II + 0,9971 x [ln (SAPS II + 1)]

e = base de los logaritmos naturales (2,718)

Fórmulas originales del APACHE II (1) y SAPS II (2)

CENTROS EN ORDEN ALFABÉTICO %

Clínica del Valle (Chubut)	0,2
Clínica Roca (Río Negro)	1,6
Clínica Suizo-Argentina (Cap. Fed.)	2,7
Hospital GA Clemente Álvarez (Cap. Fed.)	2,1
Hospital "Alejandro Khorn" (Buenos Aires)	3,1
Hospital GA "Cosme Argerich" (Cap. Fed.)	4,2
Hospital "Clemente Álvarez" (Santa Fe)	2,0

Hospital de Clínicas (Cap. Fed.)	3.6	Hospital Provincial de Neuquen	2.5
Hospital Español (Mendoza)	5.1	Hospital Regional (Chubut)	1.6
Hospital GA “Juan A. Fernández” (Cap. Fed.)	3.1	Hospital GA “Rivadavia” (Capital Federal)	1.5
Hospital “Eva Perón” (Buenos Aires)	8.2	Hospital Regional de Ushuaia (T. del Fuego)	0.7
Hospital GA “Tornú” (Cap. Fed.)	0.3	Hospital Privado del Sur FUMEB A (Bs As)	1.5
Hospital Israelita (Cap. Fed.)	3.1	Hospital GA “San Martín” La Plata (Bs As)	2.7
Hospital “Lagomaggiore” (Mendoza)	2.6	Hospital Zonal de Esquel (Chubut)	1.1
Hospital Madariaga (Misiones)	1.0	Policlínico Neuquen (Neuquen)	0.2
Hospital “Mi Pueblo” (Buenos Aires)	2.8	Policlínico Regional Villa Mercedes (San Luis)	2.5
Hospital Naval (Cap. Fed.)	2.0	Sanatorio Austral (Chubut)	0.7
Hospital GA “Paroissien” (Buenos Aires)	3.0	Sanatorio del Norte (Cap. Fed.)	0.5
Hospital “del Centenario” (Santa Fe)	6.4	Sanatorio Reconquista (Santa Fe)	3.6
Hospital Gervasio Posadas (Buenos Aires)	7.5	Sanatorio Independencia (Bs As)	4.4
Hospital GA Pirovano (Cap. Fed.)	5.5	Sanatorio Parque (Tucumán)	4.3
Hospital Italiano de Buenos Aires (Cap. Fed.)	2.1		

Provincia	Investigador 1	Investigador 2
Buenos Aires		
-HIGA Paroissien	Jose Antonio Spolidoro	Norma Rosales
-Hospital Posadas	Bernardo Maskin	Noemí Cacace Linares
-Hospital Privado del Sur “Fumeba”	Daniel Fainstein	Carlos Solá
-Clinica Independencia	Daniel Ceraso	Marcelo Rivarola
-Hospital Khorn	Patricia Martinez	Jorge Parisi
-Hospital Eva Perón	Cristina Santos	Eduardo Capparelli
-Hospital Mi Pueblo	Claudia Ceccotti	Feliciano Petro Moreno
-HIGA San Martin de La Plata	Rosa Reina	Francisco González
Capital Federal		
-HGA Pirovano	Ofelia María Miseta	Sofía Mercedes Rolando
-HGA Rivadavia	Jorge Benavente	Edgardo Slipace
-Sanatorio Del Norte	Guillermo Bordoli	Jorge Villanueva
-Clinica Suizo Argentina	Miguel Curone	Eduardo Malvino
-HGA T. Alvarez	Antonio Abdala	Sara Rost
-HGA J.A Fernández	Carlos María Dillon	María del Carmen Rozas
-HGA C.Argerich	Aldo Carusso	Graciela Cueto
-Hospital de Clínicas José de San Martin	Leopoldo Revich	Edgardo García Espina
-Hospital E.Tornú	Ariel Oscar Alvarez	Gerardo Fernández
-Hospital Israelita “Ezrah”	Alejandro Rodriguez Ricardo	Sandra Arcieri
-Hospital Naval	Patano	Sandra De Souza Bonit
Santa Fe		
-Hospital Del Centenario	Juan Figueroa Casas	Luis Pablo Cardonnet
-Sanatorio Reconquista	Francisco Martino	Gabriel Martino
-Hospital Clemente Alvarez	Juan Sanmoto	Guillermo H. García
Tucumán		
-Sanatorio Parque	José Enrique Riarte	Ricardo Wercuer
Neuquén		
-Hospital Provincial de Neuquén	Victor Sciuto	Oscar Alberto Humar
-Policlínico Neuquén	Alberto Eduardo Gian	María Alejandra Sagar

Provincia	Investigador 1	Investigador 2
Chubut -Hospital Regional -Sanatorio Austral -Clinica Del Valle -Hospital Zonal de Esquel	Eloy García Sergio Marrero Sonia Mabel Filippus Cohen Arazi	VivianaRomero Viviana Romero Daniel Ruella
Rio Negro -Clínica Roca	Miguel Angel Delgado	Cristina Orlandi
San Luis -Policlinico Regional de Villa Mercedes	Silvia Perez Betti	Francisco Criado
Tierra del Fuego -Hospital Regional Usuahia	Marcos Colman	Edgardo Solano
Mendoza -Hospital Lagomaggiore -Hospital Español	Jose Luis Marengo Walter Vazquez	Juan Sánchez Lorena Parra
Misiones -Hospital Madariaga	Miguel Angel Faraudo	

REFERENCIAS

- Schuster DP, Predicting outcome after ICU admission – The Art and Science of Assessing Risk. Chest 1992;102:1861-70.
- Hoyt JW, Leisifer DJ, Rafkin HS, Critical Care Unit. In Wenzel RP, ed. Assessing quality health care: perspectives for clinicians. Baltimore: Williams & Wilkins, 1992 :267-296
- Knaus WA, Wagner DP, Draper EA et al, The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults, Chest 1991; 100: 1619 – 1636.
- Rowan KM, Kerr JH, Major E, et al. 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 – 1401.
- Beck DH, Taylor BL, Millar B: Prediction of outcome from intensive care: A prospective cohort study comparing APACHE II and APACHE III prognostic systems in a United Kingdom intensive care unit. Crit Care Med 1997; 25:9-15.
- Angus DC: Scoring system fatigue and the search for a way forward. Crit Care Med 2000; 28:2145-2146.
- Mourouga P, Goldfrad C, Rowan K : Does it fit? Is it good? Assessment of scoring systems. Current Opinion in Critical Care 2000; 6:176-180.
- Knaus WA, Zimmerman JE, Wagner DP et al: APACHE - acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med 1981; 9 : 591-597.
- Le Gall JR, Loirat P, Alperovitch A et al: A simplified acute physiology score for ICU patients. Crit Care Med 1984; 12: 975-977.
- Knaus WA, Draper EA, Wagner DP et al: APACHE II: A severity of disease classification system. Crit Care Med 1985; 13:818-829 .
- Lemeshow S, Teres D, Pa stides H et al. A method for predicting survival and mortality of ICU patients using of objectively deriv weights. Crt Care Med. 1985; 13: 519-23.
- Lemeshow S, Teres D, Avrunin J, et al. Refining Intensive cre units outcome prediction by using changing probabilities of mortality. Crit Care Med. 1988; 16: 470-75.
- Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology score(SAPS II) based on a European / North american multicenter study. JAMA 1993; 270:2957-2963.
- Zimmerman JE, Knaus WA, Judson JA, et al. Patient selection for intensive care: A comparison of New Zealand and United States hospitals. Crit Care Med. 1988; 16: 318-326 .
- Castella X, Artigas A, Bion J, et al: A comparison of severity of illness scoring systems for intensive care unit patients:Results of a multicenter multinactional study. Crit Care Med 1995; 23:1327-1335
- Markgraf R, Deutschnoff G, Pientka L: Comparison of Apache II and III and SAPS 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
- Sirio CA, Tajimi K, Tase C et al: An initial comparison of intensive care in Japan and the United States. Crit Care Med 1992; 20:1207-1215.

18. Hosmer DW, Lemeshow S. Applied logistic regression. New York, NY: John Wiley & Sons Inc;1989.
19. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 1982; 143: 29-36 .
20. Ruttimann UE : Statistical approaches to development and validation of predictive instruments. Crit Care Clinics 1994; 10: 19-35 .
21. Teres D, Lemeshow S: As american as apple pie and APACHE. Crit Care Med 1998; 26:1297-1298 .
22. Rowan KM, Kerr JH, Major E, et al: Intensive Care Society's APACHE II study in Britain and Ireland-II: Outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. BMJ 1993; 307:977-981.
23. Rowan KM , Kerr JH , Major E , et al : Intensive Care Society's APACHE II study in Britain and Ireland-II : Variations in case-mix of adult admissions to general intensive care units and impact on outcome . BMJ 1993 ; 307:972-977
24. Moreno R, Dinis Reis Miranda, Valdivia Fider: Evaluation of two outcome prediction models on an independent database. Crit Care Med 1998; 26:50-61 .
25. Vassar MJ, Holcroft JW: The case against using the APACHE system to predict intensive care unit outcome in trauma patients. Crit Care Clinics 1994 10:117-126.
26. Teres D: Comment on "The case against using the APACHE system to predict intensive care unit outcome in trauma patients." Crit Care Clinics 1994 10:127-128
27. Watts CM, Knaus WA: Comment on "The case against using the APACHE system to predict intensive care unit outcome in trauma patients." Crit Care Clinics 1994 10:129-134.
28. Turner JS, Mudaliar YM, Chang RW et al: Acute Physiology and Chronic Health evaluation (APACHE II) scoring in a cardiothoracic intensive care unit. Crit Care Med 1991; 19:1266-1269 .
29. Katsaragakis S, Papadimitropoulos K, Antonakis P: Comparison of APACHE II and SAPS II scoring systems in a single Greek intensive care unit. Crit Care Med; 28:426-432
30. Angus DC, Clermont G, Kramer DJ : Short-term and long-term outcome prediction with the Acute Physiology and chronic Evaluation II system after orthotopic liver transplantation. Crit Care Med. 2000; 28: 150-156 .
31. Livingston BM, MacKirdy FN, Howie JC, et al: Assessment of the performance of five intensive care scoring models within a large Scottish database. Crit Care Med 2000; 28 : 1820-1827.
32. Sirio CA, Shepardson LB, Rotondi AJ, et al.: Community-wide assessment of intensive outcomes using a physiologically-based prognostic measure: Implications for critical care delivery from Cleveland Health Quality Choice. Chest 1999; 115:793-801