

Predicción de mortalidad en cuidado intensivo

Médicos, Apache II y MPM

Rodolfo Dennis, Alejandro Casas, Manuel Urina, Ada Brainsky, María Nelsy Rodríguez

Objetivo del estudio: establecer y comparar la precisión de la predicción de muerte establecida por médicos, el modelo de predicción de mortalidad (MPM) y el Apache II. **Diseño:** cohorte analítica concurrente. **Lugar:** unidad de cuidado intensivo (UCI) en un hospital de nivel terciario de Santa Fe de Bogotá, Colombia. **Pacientes:** sujetos mayores de 14 años de edad que necesitaron cuidado intensivo durante su hospitalización. **Intervenciones:** se obtuvieron estimativos para la probabilidad de muerte hechos por médicos y por ambos modelos predictivos para cada paciente en las primeras 24 horas de estancia en la UCI. **Resultados:** se recolectó información en 330 pacientes consecutivos. La mortalidad global fue de 24%. Tanto el Apache II como el MPM conservaron un poder discriminatorio, pero no tan alto cuando se comparó con su desempeño en la población norteamericana ($p < 0.02$).

Ambos grupos de clínicos tuvieron una mejor capacidad discriminatoria comparada con aquella de los modelos ($p < 0.01$ para las diferencias en áreas bajo las curvas de características operativas del receptor, COR). **Conclusiones:** el MPM y especialmente el Apache II, conservaron su poder discriminatorio y predictivo, pero sus resultados fueron menos satisfactorios en nuestro entorno; esto podría estar en relación con las diferencias en prácticas de salud en nuestro medio. Nuestros resultados apoyan su uso para la estratificación de la severidad en grupos de pacientes y para la evaluación de la prestación de los servicios de salud; su uso para la toma de decisiones en pacientes individuales en medios similares al nuestro puede no ser adecuado. Los clínicos tuvieron una mejor capacidad discriminatoria que los modelos, pero sobreestimaron consistentemente el riesgo de muerte para todos los pacientes; esto sugiere que sus predicciones solas pueden ser también insuficientes para tomar decisiones con respecto a las admisiones de UCI.

Introducción

Existe un interés creciente a nivel mundial para distribuir mejor los escasos recursos en el área de la salud; esto ha promovido la demanda no sólo de evidencia de efectividad y eficiencia, sino también de una mejor definición de quiénes se beneficiarán de este ambiente de alta tecnología que tipifica la atención actual en salud. Es por esto que la comunidad médica y la sociedad entera vuelcan su atención a las unidades de cuidado intensivo (UCI) y a decisiones tales como a quiénes se les debe restringir o retirar el tratamiento en estas unidades (1-6).

Dr. Rodolfo Dennis Verano, Msc, PhD (cand): Unidades de Cuidado Intensivo y Neumología, Departamento de Medicina Interna, Hospital Universitario de San Ignacio. Unidad de Epidemiología Clínica y Bioestadística, Facultad de Medicina, Pontificia Universidad Javeriana; Dr. Alejandro Casas: Fellow de Neumología, Departamento de Medicina Interna, Hospital Universitario de San Ignacio. Facultad de Medicina, Pontificia Universidad Javeriana. Actualmente Neumólogo, Fundación Neumológica de Colombia; Dr. Manuel Urina Triana: Unidades de Cuidado Intensivo y Cardiología. Departamento de Medicina Interna, Hospital Universitario de San Ignacio. Facultad de Medicina, Pontificia Universidad Javeriana; Dra. Ada Brainsky, Msc: Unidad de Epidemiología Clínica y Bioestadística, Facultad de Medicina, Pontificia Universidad Javeriana; Dra. María Nelsy Rodríguez, MPH: Unidad de Epidemiología Clínica y Bioestadística, Facultad de Medicina, Pontificia Universidad Javeriana, Santa Fe de Bogotá.

Debido a que ambas decisiones requieren de predicciones precisas con respecto a la probabilidad de muerte por parte del clínico a cargo, se dirigen esfuerzos encaminados a desarrollar un método objetivo de estimar dicha probabilidad. La necesidad de este estimativo objetivo encuentra su apoyo en el hecho de que las decisiones médicas pueden ser subjetivas (6) y dadas a sesgos (7, 8), y en que los clínicos no calculan el pronóstico adecuadamente, llevando posiblemente a una falta de equidad en sus estimativos (9).

Dado el relativamente reciente desarrollo de estrategias estadísticas con implicaciones claras no sólo con respecto a pronóstico y predicción (10-13), sino también a estándares aceptados para desarrollo y aplicación (14), existe una tendencia reciente a introducir y evaluar estas estrategias en el proceso de toma de decisiones, especialmente la decisión de restringir o suspender tratamientos en la UCI (2, 6).

Existen por lo menos dos modelos multivariados, el Apache II y el MPM (15, 16), desarrollados para la estimación de la probabilidad de muerte en grupos de pacientes admitidos a la UCI, que hemos usado para comparar unidades de cuidado intensivo en nuestro medio (17). Estos modelos han sido cada vez más estudiados para tratar de tomar decisiones en pacientes individuales (1, 2, 6, 8). Sin embargo, antes de que estos u otros modelos, estáticos o dinámicos a través del tiempo, sean aceptados para este propósito, deben ser validados en una gran variedad de sitios y comparados localmente con las predicciones clínicas. Comparaciones recientes de las

reglas de predicción y estimativos de mortalidad hechos por clínicos han arrojado diferentes resultados (8, 18-21): ni el MPM ni el Apache II han sido validados adecuadamente en hospitales universitarios de nivel terciario en países en vía de desarrollo. Es un hecho conocido que la variación geográfica de la presentación de las enfermedades o las diferencias en los tratamientos pueden alterar las propiedades "estables" de los exámenes diagnósticos (22) y que las reglas de predicción clínica pueden cambiar dependiendo de la prevalencia local de la enfermedad (11, 23).

En conclusión, aunque una evaluación válida de riesgo a través de modelos multivariados puede ofrecer una ayuda al proceso de toma de decisiones en las unidades de cuidado intensivo en todo el mundo (especialmente para aquellas en países en vía de desarrollo, donde los recursos son escasos), estos modelos deben ser validados localmente y comparados con los estimativos hechos por los clínicos. Esto es particularmente deseable en unidades de cuidado intensivo con características o tasas de mortalidad diferentes de aquéllas donde los modelos fueron validados (24), teniendo en cuenta que éstos podrían ser utilizados para toma de decisiones que generalmente involucran situaciones de vida o muerte.

El hecho de que el MPM y el Apache II no llenen las expectativas de validación, implicaría que estos modelos y algunos otros usados de rutina (Glasgow, Ramson) tendrían que ser utilizados con algunas precauciones, o ni siquiera deberían ser implementados. Esto impulsaría a los investigadores de países en vía

de desarrollo a crear modelos ajustados a sus patrones geográficos de enfermedad.

Material y Métodos
Diseñamos un estudio de cohortes que se desarrolló en la UCI del Hospital San Ignacio de Santa Fe de Bogotá, Colombia. Utilizamos la información de todos los pacientes que ingresaron a la UCI. El Hospital San Ignacio (HSI) es un hospital universitario, centro de referencia de nivel terciario, con 300 camas. La población que se sirve de esta institución consiste en una mezcla de pacientes del área urbana de Bogotá y de zonas rurales periurbanas.

La UCI del HSI es una unidad médico-quirúrgica de 11 camas con un alto índice de ocupación. Ofrece alta tecnología para estándares colombianos; tiene capacidad para el uso de rutina de ventiladores, hemodiálisis y diálisis peritoneal, cateterización de la arteria pulmonar y monitoreo intraarterial. La unidad cuenta con un director de tiempo completo. Tanto un cardiólogo como un neumólogo están de turno de disponibilidad, que se va cambiando semanalmente. Residentes de segundo y tercer año de medicina interna y anestesiología hacen rotaciones de tres meses e intervienen en el cuidado de los pacientes. La relación enfermera-paciente es de 1:2 para los turnos diurnos y de 1:3 en los turnos nocturnos.

A. Criterios de exclusión

Los criterios de exclusión fueron extraídos casi sin variaciones del trabajo original de Knaus et al y Teres et al (15, 16). Fueron excluidos los pacientes quemados, los menores de 14 años,

los que tengan información incompleta de las variables en estudio, los sometidos a cirugía cardíaca o con necesidad de cuidado intensivo coronario y los pacientes que murieran en las seis horas posteriores a su ingreso a la UCI.

También excluimos los pacientes remitidos a otras UCI en hospitales diferentes, o transferidos a otro hospital después de haber salido de la UCI; esto con el fin de disminuir la posibilidad de introducir sesgos de tratamiento, a pesar de que no acostumbramos a transferir sistemáticamente pacientes con un diagnóstico determinado o con un grado de severidad establecido, de nuestra UCI a otro hospital.

B. Tamaño de la muestra

No existe un consenso en la literatura acerca de la mejor manera de calcular el tamaño de la muestra cuando se trata de desarrollar o validar un modelo multivariado. Harrell (13) recomienda no intentar un análisis de regresión para este propósito cuando existen menos de diez veces el número de eventos en la muestra como variables potenciales de predicción. Wasson (14) afirma que debe haber por lo menos cinco pacientes con el desenlace de interés para cada variable predictiva en la ecuación. La ecuación predictiva en el Apache II tiene tres variables en el modelo, por lo que se necesitan entre 15 y 30 muertes intrahospitalarias; el MPM tiene 11 variables requiriéndose entonces entre 50 y 110 de estos desenlaces. Nosotros recolectamos la información en pacientes consecutivos hasta que se registraron 70 muertes intrahospitalarias.

C. Definición de las variables

Las variables fueron recolectadas mediante cuestionarios estandarizados. Para el MPM y el Apache II la recolección de variables fue conducida en el mismo formato de los trabajos originales. El doctor Lemeshow nos envió los esquemas originales de la recolección de datos del MPM. El formato del Apache II era una réplica exacta del utilizado por Knaus et al en su trabajo (15).

D. Métodos

Para la validación del MPM y del Apache II los pacientes fueron identificados en el momento de la admisión a la UCI por dos residentes entrenados en recolección de información en la Unidad de Epidemiología Clínica de la Facultad de Medicina. Todos los días uno de los residentes examinaba las nuevas admisiones de la UCI y se cercioraba de que se respetaran los criterios de exclusión.

Si no había duda en cuanto a los criterios de exclusión, el residente recolectaba la información de los datos demográficos. Si existía alguna duda con respecto al estado del paciente, el investigador principal o uno de los coautores definía si el paciente podía ser o no incluido, o si se requería de un panel para consenso antes de ser incluido en el estudio. Este proceso se llevó a cabo sin conocer la evolución del paciente.

Después de recolectada la información, el residente seguía al paciente para documentar evidencias de variables de desenlace, o hasta que el paciente dejaba la UCI, momento en el cual nuevamente verificaba los diagnósticos de salida de la UCI, bus-

cando una vez más criterios de exclusión. Cada residente entonces seguía al paciente hasta cuando era enviado a otra institución, hasta que se daba de alta a la casa (las unidades de cuidado intermedio se consideraban como hogar) o hasta que moría en el hospital.

Después de dar la salida, un asistente de investigación revisaba las historias clínicas y recolectaba las variables necesarias para el MPM y el Apache II. La información para el MPM se recolectaba al momento de la admisión y para el Apache II se hacía después de las primeras 24 horas de estancia en la UCI. La confiabilidad intraobservador (asistente de investigación) se comprobaba al revisar nuevamente las historias clínicas antes revisadas por éste (no menos de 2 semanas). Fueron utilizados en esta evaluación los métodos usuales de concordancia (25).

El establecimiento de probabilidad de mortalidad por parte de residentes e instructores se hacía tan pronto como se aprobaba una nueva admisión a la UCI o tan pronto como el paciente llegaba a la UCI. El residente de turno en la UCI escribía su predicción de muerte y posteriormente le pedía al instructor de turno su respectiva predicción, registrando independientemente ambos estimativos. Así teníamos para cada paciente una predicción hecha por el residente y otra por el profesor de turno correspondiente. Ninguno de ellos conocía la probabilidad asignada por el MPM o por el APACHE, y se les recordaba continuamente el calcular dicha probabilidad individualmente.

Aunque esperábamos que la mayoría de los estimativos se dieran durante la admisión, algunos fueron difíciles de obtener en el momento del ingreso de los pacientes, debido a consideraciones logísticas. Incluimos todos los estimativos de probabilidad hechos en las primeras 12 horas de estancia en la UCI. La probabilidad de muerte del MPM y del Apache II fue calculada usando una hoja electrónica (Lotus 1,2,3), como lo sugieren Lemeshow et al (26). No se incluyeron métodos de identificación personal en los archivos de análisis.

E. Análisis estadístico

La información fue almacenada en un microcomputador y la mayoría de los análisis fueron llevados a cabo por paquetes de BMDP (*biomedical programs*) (27). Primero se utilizaron medidas descriptivas, caracterizando muertes y sobrevivientes, basándose en los factores (exposiciones) que se creía tenían alguna relación con el riesgo de muerte; éstos incluían todas las variables independientes necesarias para el cálculo del Apache II y del MPM. Se emplearon métodos convencionales para las sinopsis (promedios, medianas y porcentajes) de las comparaciones y pruebas de significancia, tales como la t de Student, el Chi cuadrado o la prueba exacta de Fisher cuando fue necesario.

Se construyeron curvas de características operativas del receptor (COR) comparando la tasa de verdaderos positivos contra la de falsos positivos a diferentes puntos de corte, con el fin de obtener una perspectiva visual de las características operativas del Apache II, del MPM y de los

instructores y residentes a través de una amplia gama de probabilidades y posteriormente para determinar el poder discriminatorio al comparar las áreas bajo la curva, como sugieren Hanley y McNeil (28, 29).

Las pruebas de bondad del ajuste empleadas en la validación de estos modelos para determinar su precisión predictiva en nuestro medio, fueron aquellas desarrolladas por Hosmer y Lemeshow (30). Estas pruebas computan el número esperado de muertes al agrupar a los pacientes dentro de diez diferentes estratos definidos por su probabilidad estimada de mortalidad. La comparación entre los números observados y los esperados en cada estrato permite una determinación del ajuste del modelo predictivo. Las estadísticas resultantes (H_g, C) son entonces comparadas con una distribución de tipo Chi cuadrado para determinar así la significancia, con 10 grados de libertad, ya que se trataba de un modelo desarrollado previamente (30). También comparamos la tasa total de clasificaciones correctas para cada punto de corte. Lo mismo se hizo para las predicciones hechas por residentes e instructores.

Resultados

La información fue recolectada en 546 admisiones consecutivas a la UCI, en un período de 21 meses. Doscientos siete pacientes tenían criterios de exclusión. Generalmente se trataba de diagnósticos de enfermedad coronaria, trastornos relacionados con ésta, o por muerte temprana en la UCI.

Trescientos treinta pacientes se incluyeron en el estudio. La edad promedio fue de 52 años; 32%

de los sujetos eran mayores de 65 años. La razón entre hombres y mujeres fue de 0.85. El porcentaje de pacientes admitidos con enfermedades crónicas y cáncer fue de 22 y 4, respectivamente. La mortalidad global para este grupo fue de 24%.

La información para el MPM pudo ser obtenida en 98% de los casos y para el Apache II en 90% (debido a historias no disponibles o información incompleta). La predicción de muerte hecha por residentes y por el personal de turno de disponibilidad en las primeras 12 horas de estancia en la UCI pudo ser obtenida en 82 y en 78% de los casos respectivamente. Para las comparaciones estadísticas entre los modelos y los clínicos, empleamos pacientes con documentación completa de todas las variables en estudio.

A. Validación del Apache II y el MPM

Determinamos la capacidad discriminatoria de ambos modelos (MPM y Apache II) ajustando las curvas COR a nuestra información. Como puede verse en la Figura 1 y en la Tabla 1, tanto el MPM como el Apache II pierden poder en nuestro medio en forma significativa, al compararse con las características operativas de éstos en los Estados Unidos, de donde los cálculos pudieron ser extraídos y analizados (15, 16). No hubo diferencias estadísticamente significativas en nuestro medio entre los dos modelos predictivos con respecto al área bajo las curvas COR ($p > 0.20$).

La pérdida de poder discriminatorio al parecer surgió especialmente de una disminución en la sensibilidad (capacidad de

Mortalidad en cuidado intensivo

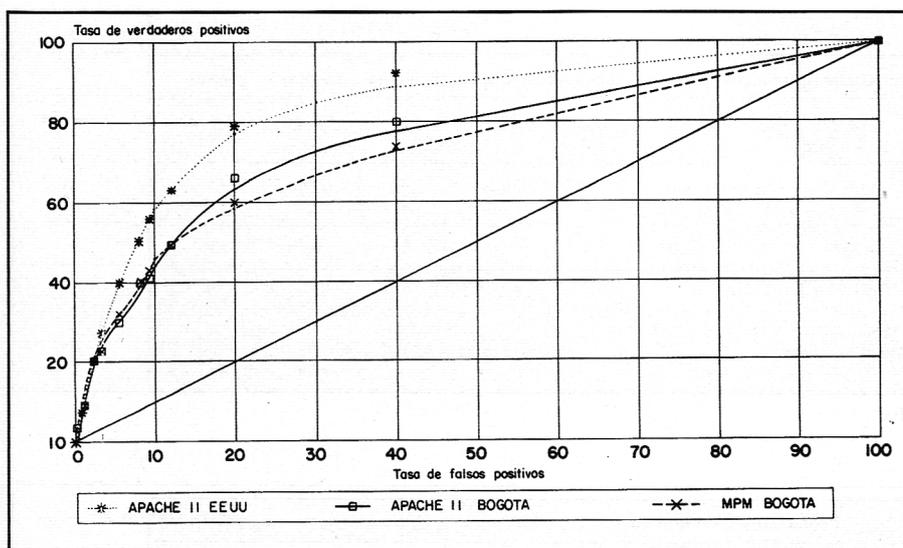


Figura 1. Curvas de características operativas del receptor (COR) del Apache II en los EE. UU. y del MPM y Apache II en nuestro medio. El área debajo de la curva es una medida de la capacidad discriminatoria de predicción, independientemente del punto de corte escogido.

	EEUU	Bogotá	Valor de P
MPM	0.86	0.73	0.0001
Apache II	0.86	0.77	<0.02
Residentes vs Instructores	-	0.88 0.84	<0.19
Residentes vs MPM	-	0.88 0.73	<0.001
Residentes vs Apache II	-	0.88 0.77	<0.007

(*) Características operativas del receptor

Tabla 1. Areas bajo la curva COR (*) para Apache II, MPM y clínicos

Variable	M/A	Muerto	Vivo	Valor de P
Edad en años (*)	M/A	53.6 (20)	51.6 (19)	0.07
Presión arterial sistólica	M	103 (41)	131 (37)	0.00
Frecuencia cardiaca	M-A	104 (29)	93 (27)	0.00
APS (**)	A	14.3 (7)	6.9 (5)	0.00
Enf. crónica (%)	A	35.1	17.3	0.01
Paciente quirúrgico (%)	M	33.8	43.8	0.35
Enfermedad renal (%)	M-A	10.1	5.3	0.00
Ingreso previo a UCI	M	12.7	10.1	0.20
RCP (***) (%)	M	11.4	4.5	0.00
Cáncer (%)	M-A	0.9	2.02	0.00
Infección (%)	M	64.6	36.0	0.00
Coma (%)	M-A	32.9	18.6	0.02

(*) Media
 (**) Puntaje fisiológico agudo
 (***) Resucitación cardiopulmonar
 M= MPM A= Apache II

Tabla 2. Análisis de variables usadas para predicción de mortalidad por el MPM y el Apache II.

detectar muertes). Sin embargo, el Apache II tuvo una capacidad discriminatoria adecuada en el punto de corte para probabilidad de muerte de 0.40 con una sensibilidad y una especificidad de 40% y de 96% respectivamente. La clasificación correcta en general para este punto de corte fue de 83%. Para el MPM, la mejor de las clasificaciones correctas fue de 80%, que ocurrió en el punto de corte de 0.50 para probabilidad de muerte.

Cuando las variables individuales usadas para el MPM y el Apache II fueron aplicadas en nuestro medio, algunas no se asociaron con muerte en el análisis univariado ni en el multivariado (regresión logística). La Tabla 2 muestra algunas de estas variables y su valor de P correspondiente obtenido a través de regresión logística.

Para establecer la capacidad predictiva de ambos modelos, comparamos el número esperado de muertes y de sobrevivientes como lo sugieren Hosmer y Lemeshow. Las Tablas 3 y 4 muestran este análisis. Los dos modelos tienen diferencias significativas entre lo esperado y lo observado dentro de cada estrato, ambos con la estadística C o Hg (los resultados se muestran para la Hg). Sin embargo, mientras que el MPM subestimó la probabilidad de muerte en los estratos de bajo riesgo y la sobreestimó en los de alto riesgo, el Apache II sistemáticamente subestimó la tasa observada de muertes, de tal manera, que la tasa de muertes observada versus la esperada, usando el Apache II, fue de 1.4. El ajuste obtenido con el Apache II puede ser un poco mejor (como se ilustra para el valor de la estadística Hg);

Probabilidad de muerte (*)											
	≤ 0.09	0.10-0.19	0.20-0.29	0.30-0.39	0.40-0.49	0.50-0.59	0.60-0.69	0.70-0.79	0.80-0.89	0.90-1.0	Total
Muerto	6.0 19	8.86 9	7.93 10	6.2 5	61.6 7	5.4 6	2.5 3	3.76 2	6.81 4	14.32 13	78
Vivo	148.0 135	53.14 53	25.07 23	11.8 13	7.84 7	4.6 4	1.5 1	1.24 3	1.19 4	0.86 2	245
	154	62	33	18	14	10	4	5	8	15	323
(**) Hg: 44.8 P 0.001 $\Sigma O/\Sigma e = 78/67.94 = 1.14$ (***) (*) En deciles de riesgo (**) e = esperado 0 = observado (***) $\Sigma O/\Sigma e =$ razón de lo observado vs lo esperado (Lo ideal es 1)											

Tabla 3. Bondad del ajuste de los estimativos del MPM.

Probabilidad de muerte (*)											
	≤ 0.09	0.10-0.19	0.20-0.29	0.30-0.39	0.40-0.49	0.50-0.59	0.60-0.69	0.70-0.79	0.80-0.89	0.90-1.0	Total
Muerto	615. 14	7.97 9	6.7 12	7.13 8	5.34 8	4.93 6	3.85 5	4.42 5	1.72 2	0.9 1	71
Vivo	148.85 141	49.03 48	21.3 16	13.87 13	6.66 4	4.07 3	2.15 1	1.58 0	0.28 0	0.10 0	226
	155	57	28	21	12	9	6	6	2	1	297
(**) Hg: 22.7 P 0.002 $\Sigma O/\Sigma e = 1.40$ (***) (*) En deciles de riesgo (**) e = esperado 0 = observado (***) $\Sigma O/\Sigma e =$ razón de lo observado vs lo esperado (Lo ideal es 1)											

Tabla 4. Bondad del ajuste de los estimativos del Apache II.

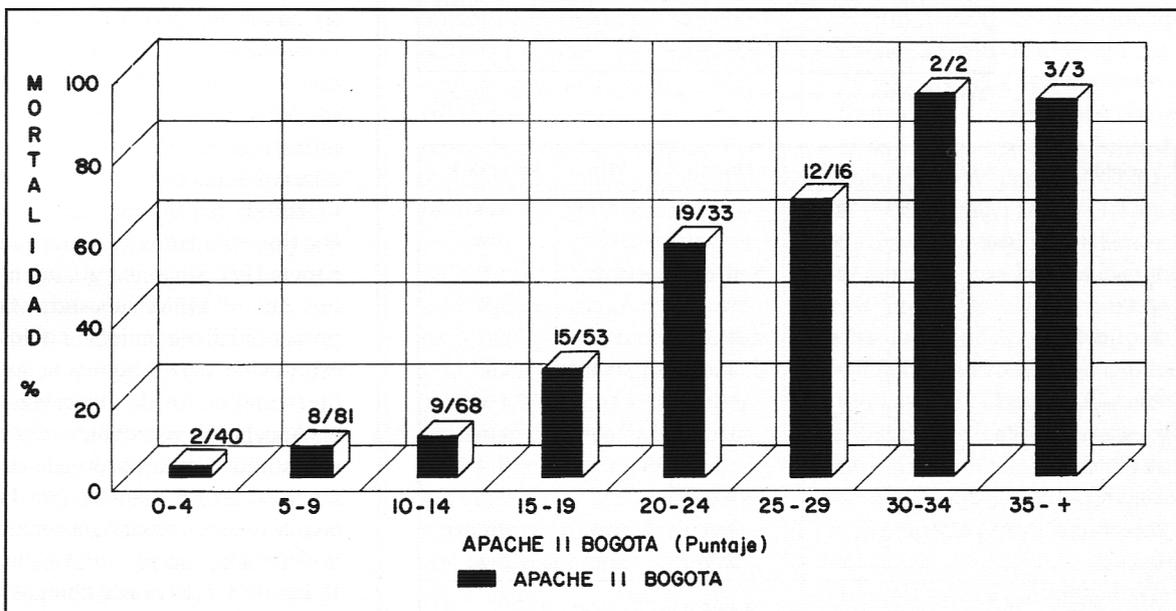


Figura 2. El puntaje de severidad de enfermedad del Apache II muestra una tendencia significativa con respecto a la mortalidad hospitalaria en nuestro medio como ha sido documentado previamente para grupos de pacientes. Ningún paciente con un puntaje de Apache II de 30 o más sobrevivió en nuestro hospital, pero los pacientes clasificados dentro de este estrato eran pocos (5/5).

Mortalidad en cuidado intensivo

	Probabilidad de muerte (*)										Total
	≤ 0.09	0.10-0.19	0.20-0.29	0.30-0.39	0.40-0.49	0.50-0.59	0.60-0.69	0.70-0.79	0.80-0.89	0.90-1.0	
Muerto	2.01 1	5.6 2	6.75 4	10.65 1	6 1	10.55 2	10.8 9	6.3 5	11.35 10	29.3 25	60
Vivo	42.99 44	41.4 47	26.25 29	24.35 34	9 14	10.45 19	7.2 9	2.7 4	2.65 4	1.87 6	210
	45	49	33	35	15	21	18	9	14	31	270

(**) Hg: 50.6 P 0.001 $\Sigma O/\Sigma e = 60/99.14 = 0.61$ (***)
 (*) En deciles de riesgo
 (**) e = esperado O = observado
 (***) $\Sigma O/\Sigma e$ = razón de lo observado vs lo esperado (Lo ideal es 1)

Tabla 5. Bondad del ajuste de los estimativos clínicos.

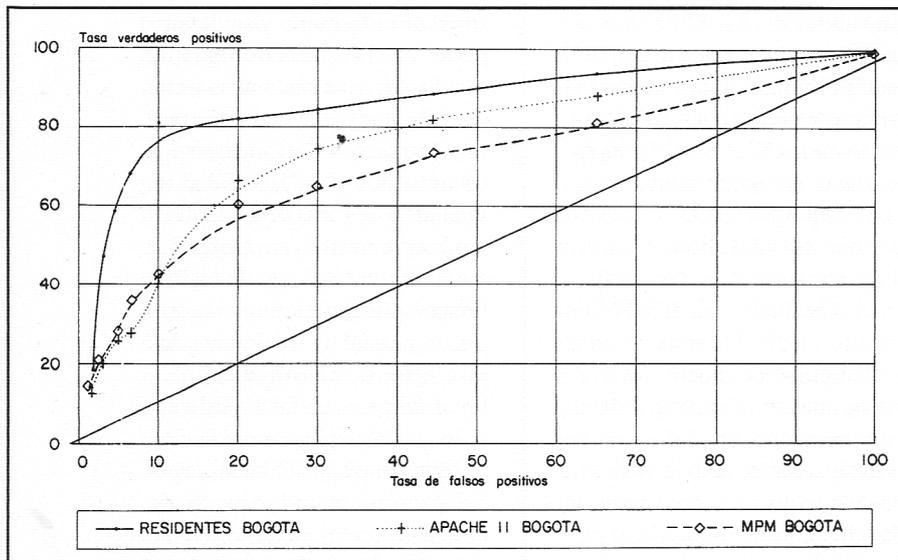


Figura 3. Curvas COR comparando las predicciones de muerte de residentes con respecto a aquellas hechas por ambos modelos en nuestro medio. Los residentes tuvieron mejor capacidad discriminatoria para cualquiera de los puntos de corte escogidos; ellos tuvieron la capacidad de hacer predicciones de muerte más acertadas con una tasa de falsos positivos más baja.

además, el puntaje del APACHE mostró una clara y fuerte tendencia con respecto a la mortalidad hospitalaria (Figura 2).

B. Capacidad predictiva y discriminatoria de instructores y residentes de UCI

Los estimativos de residentes e instructores fueron considerados como reglas de predicción clínica. Como tales, el poder discriminatorio de ambos fue significativamente mejor que el del

MPM o el del Apache II en nuestro medio y muy similar al de ambos modelos cuando se describieron en EE.UU. (Figura 3 y Tabla 1). Sin embargo, la capacidad predictiva fue inadecuada, ya que consistentemente sobreestimaban la probabilidad de muerte en cada estrato de riesgo; de ahí que la tasa de las muertes observadas sobre las esperadas utilizando los estimativos de residentes y profesores fue de 0.61 (Tabla 5).

Discusión

Es sumamente deseable tener una estimación objetiva de pronóstico para los pacientes admitidos a la UCI. Nuestro estudio mostró que dos modelos multivariados sugeridos en la literatura como ayudas posibles en la toma de decisiones en la UCI. conservan su capacidad predictiva en nuestro hospital.

Sin embargo, a pesar de que ambos modelos conservan una capacidad predictiva y un poder discriminatorio (especialmente el Apache II), muestran una disminución importante de esta característica cuando son extrapolados a nuestro entorno; en consecuencia, el ajuste de sus predicciones (observadas versus esperadas) no es tan bueno como en otros estudios.

Las diferencias que encontramos con respecto a la capacidad predictiva entre los dos modelos puede ser debida a varios factores; entre otros, que el Apache II utiliza estimativos de las variables recolectadas en las 24 horas iniciales de ingreso a la UCI, mientras que el MPM emplea la información recolectada en el momento de admisión a la UCI; esto podría explicar, por lo menos parcialmente, los resultados observados.

Se han publicado algunos estudios de probabilidad de muerte estimada por clínicos y comparada con tasas reales de mortalidad o con una escala de severidad de enfermedad (8, 18, 32). En el estudio de Detsky et al (32), los clínicos sobreestimaron el riesgo de muerte de los pacientes en el momento de admisión a la UCI, como ocurrió en nuestro estudio. Sin embargo, no se realizó una comparación concomitante con una escala de severidad de enfermedad.

Recientemente, Krauss et al encontraron que en una UCI de los EE.UU. no había diferencias del poder discriminatorio entre el Apache II y los clínicos. A pesar de que no se consideró la posibilidad de un error tipo II y de que no se llevaron a cabo las bondades de ajuste estadísticas, sus hallazgos sugieren que el proceso de toma de decisiones en la UCI podría beneficiarse del uso del Apache II, ya que éste es considerado como un método "objetivo" y equivalente al estimativo del juicio clínico (19). Resultados similares fueron obtenidos por Chang et al (8), quienes utilizaron un modelo dinámico, diario, derivado del Apache II, para calcular la probabilidad de muerte.

Nuestros resultados se encuentran más en concordancia con los Marks et al (20) y Brannen et al (21); aunque el MPM no fue evaluado en sus estudios, ellos también encontraron mejores resultados con el juicio clínico "subjetivo" que con el Apache II. En nuestro estudio los residentes parecieron tener una mejor capacidad predictiva que los instructores; este hallazgo podría deberse al hecho de que los residentes proporcionan el cuidado

de primera línea pero no son los directamente responsables de los pacientes, y por ende su estimativo puede ser más objetivo. Adicionalmente, los instructores pueden ser más conservadores cuando expresan verbalmente la probabilidad de muerte de un paciente. Finalmente, los residentes pudieron haber precisado mejor su predicción, una vez conocido el estimativo hecho por el instructor, dado que eran ellos los encargados de la recolección de ambos.

La validez del MPM ha sido estudiada previamente en una UCI en las Filipinas, donde la bondad del ajuste no fue tan satisfactoria como en los EE.UU. (31). Es interesante especular acerca de las razones por las cuales hay pérdida parcial del poder, tanto discriminatorio como predictivo, en el Apache II y en el MPM en nuestro medio. Diferencias en la probabilidad de muerte (más altas en nuestro medio), que de alguna manera no son reflejadas en la evaluación de ambos modelos, pueden explicar parcialmente la disminución en la capacidad predictiva, pero no la disminución de poder discriminatorio, ya que este análisis es independiente de la prevalencia (22). Estudios recientes han sugerido que por lo menos la fuente de admisión (unidad o piso del hospital o sala de urgencias), o el tipo de paciente (traumatizado o quirúrgico), pueden predecir la severidad de la enfermedad independientemente del puntaje del Apache II y que debido a esto, dichas situaciones pueden no estar correctamente representadas en esta escala (33-35).

Tanto el MPM como el Apache II pueden haber tenido un desempeño menos bueno en nues-

tro medio si la calidad del cuidado y de los tratamientos posteriores a la evaluación fue marcadamente diferente de aquella donde los modelos fueron desarrollados. En un trabajo previo de Knauss et al, el APACHE no fue un predictor adecuado de muerte en dos centros de los EE.UU.; en uno por subestimación y en el otro por sobreestimación del riesgo de muerte observado. Esta situación pareció correlacionarse con la calidad del cuidado en la UCI -especialmente con el personal de enfermería y su desempeño en las decisiones que involucran a los pacientes- y con las relaciones médico-enfermera. Grandes variaciones en los tratamientos y en la calidad de vida después de dejar la UCI también contribuyen a explicar estas diferencias (36). Estos hallazgos pueden ciertamente explicar parcial o totalmente los resultados en nuestro medio.

Finalmente, es un hecho bien sabido que las reglas de predicción pueden "encogerse" en su capacidad discriminatoria cuando son utilizadas en un medio diferente de aquél donde fueron desarrolladas, algunas veces incluso dentro de la misma área geográfica; estos hallazgos pueden no ser el reflejo de estrategias inadecuadas de desarrollo del modelo, de variables evaluadas subjetivamente, de diferencias en los tratamientos o de la probabilidad previa al examen de que el evento sea predicho (37,38). Esto simplemente puede indicar que los modelos multivariados empleados para modelos predictivos no son universales, ni tienen razón de serlo. Esto también puede impulsar a los clínicos e investigadores de países en desarrollo a calibrar o a adaptar escalas pre-

dictivas que se ajusten con su situación particular.

En conclusión, en nuestro estudio se mostró que el MPM y el Apache II conservan su poder discriminatorio, pero que la bondad del ajuste de ambos modelos no es tan impactante como la informada en otros estudios, pues los clínicos tienen una mejor capacidad discriminativa. Sin embargo, tanto los clínicos en entrenamiento como los instructores tendieron a sobreestimar la probabilidad de muerte para pacientes de bajo y de alto riesgo, sugiriendo quizás que tampoco sus predicciones solas pueden ser adecuadas para tomar decisiones con respecto al ingreso a la UCI. Esto refuerza la idea de que existirá un papel en el futuro para las estrategias de predicción que ayuden a la toma de decisiones en pacientes individuales de las UCI (1); por lo tanto, se requiere un mayor refinamiento de ambos modelos en evaluaciones prospectivas

Nuestro estudio apoya la hipótesis de que los modelos predictivos pueden no ser universales en su aplicación y que la validación prospectiva y la eventual calibración de la escala debe realizarse, especialmente si la toma de decisiones en pacientes individuales está siendo considerada como una herramienta útil de predicción.

Summary

Study objective: to establish and compare the accuracy of a death prediction probability in the ICU setting by physicians in training, medical attending staff, Apache II and the mortality prediction model (MPM).

Design: analytic concurrent cohort.

Setting: combined surgical and medical intensive care unit in a tertiary care facility in Santa Fe de Bogotá, Colombia.

Patients: persons over 14 years of age in need of ICU care sometime during their hospital stay, excluding those with burns, coronary care or cardiac surgery. Interventions: during the first 24 hours of ICU stay, we obtained a death probability estimate from physicians and from predictive models for each patient; we then followed all patients until their survival status was documented at hospital discharge. The comparison among predictions, as well as discriminative power and cutoff points was assessed with receptor operative characteristics (ROC) curves. The goodness of fit between observed/expected death rate was done with the Hosmer and Lemeshow statistic.

Results: we collected information on 330 consecutive patients that did not meet exclusion criteria. Mean age for the cohort was 52 years; patients with chronic disease and cancer accounted for 22% and 4% of the group respectively. Overall hospital mortality was 24%. Our results showed that both Apache II and the MPM retained their discriminative properties, but their performance was not as good as that of similar studies in the US ($p < 0.02$). Apache II consistently underestimated death probability in our setting (O/E death ratio 1.4). Both groups of physicians had better discriminative ability than the models ($p < 0.01$ for differences in area under the ROC curve). However, physicians consistently overestimated death probability (O/E death ratio 0.60).

Conclusions: the MPM, and especially Apache II, retained their discriminatory and predictive power, but less satisfactorily than in other settings; this may be related with differences in health practices in our setting as compared with settings where models were developed. Our results support their use for stratification of severity across ICU patient groups and health care delivery assessment. However, our results also suggest that their use for individual decision making in settings similar to ours should be conducted with caution. Clinicians involved in the ICU environment showed better discriminative ability, but they consistently overestimated the death risk for all patients. This also suggests that their informed predictions alone may be insufficient for individual decisions regarding admissions to the ICU.

Agradecimientos

Los autores agradecen a todos los médicos en la UCI del Hospital San Ignacio que cooperaron con este estudio, y a los doctores Sandra Norman, JA Grisso MD y Daniel Teres MD, por su revisión cuidadosa de este manuscrito y sus útiles sugerencias. Agradecemos además a la Dra. Alejandra Sanin por su ayuda en la revisión y elaboración del manuscrito. Este estudio fue posible en parte gracias a la ayuda de la fundación Rockefeller y a Inclen Inc.

Referencias

1. Chang WS. Individual outcome prediction model for ICU patients. *Lancet* 1989; **II**: 143-146.
2. Kalb PE, Miller DH. Utilization strategies for intensive care units. *JAMA* 1989; **216**: 2369-2395.
3. Gibbs M. Love and let die. *Time Magazine* 1990; March 19: 32-38.
4. Zimmerman JE, Knaus WA, Sharpe SM, et al. The use and implications of do not resuscitate orders in intensive care units. *JAMA* 1986; **235**: 351-356.
5. Bone RC. Outcomes in critical care medicine. *JAMA* 1988; **260**: 3487.
6. Jennet B. Inappropriate use of intensive care. *Brit Med J* 1984; **289**: 1709-1711.
7. Tversky A, Kahneman D. Judgement under uncertainty: Heuristics and Biases. *Science* 1974; **185**: 1124-1131.

8. **Chang RWS, Lee B, Jacobs S, et al.** Accuracy of decisions to withdraw therapy in critical ICU patients: Clinical Judgement vs. a computer model. *Critical Care Med* 1989; **17**: 1091-1097.
9. **Wachter RM, Luce SM, Hearst N, et al.** Decisions about resuscitation: Inequities among patients with different diseases but with similar prognosis. *Ann Intern Med* 1989; **111**: 525-532.
10. **Cook EF, Goldman L.** Empiric comparison of multivariate analytic techniques. *J Chronic Dis* 1984; **37**: 721-731.
11. **Goldman L, Cook EF, Brand DA, et al.** A computer protocol to predict myocardial infarction in emergency department patients with chest pain. *N Engl J Med* 1988; **318(13)**: 797-803.
12. **Harrell FE, Lee KL.** Regression modeling strategies for improved prognostic prediction. *Statistics in Medicine* 1984; **3**: 143-152.
13. **Harrell FE, Lee KL, Matchar DB, et al.** Regression models for prognostic prediction. *Cancer treatment reports* 1985; **69(10)**: 1071-1077.
14. **Wasson JH, Sox HC, Neff RK, et al.** Clinical prediction rules: applications and methodological standards. *N Engl J Med* 1985; **313(13)**: 793-799.
15. **Knaus WA, Draper EA, et al.** Apache II: a severity of disease classification system. *Critical Care* 1985; **13(10)**: 818-829.
16. **Lemeshow S, Teres D, Spitz J, et al.** Refining intensive care unit outcome prediction by using changing probabilities of mortality. *Crit Care Med* 1988; **16(5)**: 470-477.
17. **Dennis R, Acero R, Salas C, Orejuela F.** Evaluación del Cuidado Intensivo. *Acta Med Colomb* 1995; **20**: 64-70.
18. **Krusse JA, Thill MC, Carlson RW.** Comparison of clinical assessment with Apache II for predicting mortality risk in patients admitted to a medical intensive care unit. *JAMA* 1988; **260**: 1739-1742.
19. **Silverstein MD.** Prediction instruments and clinical judgement in critical care. *JAMA* 1988; **260**: 1758-1759.
20. **Marks RJ, Simons RS, Blizzard RA, Browne DR.** Predicting outcome in intensive therapy units. *Intensive Care Med* 1991; **17(3)**: 159-163.
21. **Brannen AL, Godfrey LJ, Goetter WE.** *Arch Intern Med* 1989; **149(5)**: 1083-1086.
22. **Poses RM, Cebul RD, Collins M.** The importance of disease prevalence in transporting clinical prediction rules. *Ann Intern Med* 1986; **105**: 586-591.
23. **Dombal FT, Staniland JR, Clamp SE.** Geographical variation in disease presentation. *Med Decis Making* 1981; **1(1)**: 59-69.
24. **Dragsted L, Jorgensen J, Niels-Hennrik J, et al.** Interhospital comparisons of patient outcome from intensive care: importance of lead-time bias. *Crit Care Med* 1989; **17**: 418-422.
25. **Kramer MS, Feinstein AR.** The Biostatistics of Concordance. *Clin. Pharmacol. Ther.* 1981; **29(1)**: 111-122.
26. **Lemeshow S, Pastides H, Avrunin JS, Teres D.** Epidemiologic Programs for Computers and Calculators. *Am J Epidemiol* 1985; **122(4)**: 710-716.
27. **BMDP Statistical Software.** University of California, 1985.
28. **Mcneil B J, Adelstein S J.** Determining the value of diagnostic and screening tests. *J Nucl Med* 1976; **17(6)**: 439.
29. **Hanley JA, Macneil BJ.** The meaning and use of the area under the receiver operating characteristics (ROC) curve. *Diagnostic Radiology* 1982; **143**: 29-36.
30. **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(1)**: 92-106.
31. **Beltran AM, et al.** "Goodness of fit" of the Mortality Prediction Model at the Philippine General Hospital. Abstract, Seventh Annual Meeting of the International Clinical Epidemiology Network. January, 1989.
32. **Detsky AS, Stricken SC, Mulley AG, et al.** Prognosis, survival, and the expenditure of hospital resources for patients in an intensive care unit. *N Engl J Med* 1981; **305**: 667-672.
33. **Escarce JJ, Kelley MA.** Admission source to the medical intensive care unit predicts hospital death independent of Apache II score. *JAMA* 1990; **264**: 2389-2394.
34. **Civetta JM, Nelson LA, Hudson-Civetta JA.** Evaluation of Apache II for cost-containment and quality analysis in surgical patients. *Ann Surg* **212**: 266, 1990.
35. **Cerra FB, Negro F, Abrams J.** Apache II score does not predict MOF or mortality in postoperative surgical patients. *Arch Surg* 1990; **125**: 519-522.
36. **Knaus W, Draper EA, Wagner DP, et al.** An evaluation of outcome from intensive care in mayor medical centers. *Ann Intern Med* 1986; **104**: 410-418.
37. **Rose CC, Murphy JG, Schwartz JS.** Performance of an index predicting the response of patients with acute bronchial asthma to intensive emergency department treatment. *N Engl J Med* 1984; **310**: 573-577.
38. **Centor RM, Yarbrough B, Wood JP.** Inability to predict relapses in acute asthma. *N Engl J Med* 1984; **310**: 577-580.