

Validación y comparación de tres puntuaciones de predicción de mortalidad por neumonía adquirida en la comunidad, en pacientes del servicio de urgencias

Validation and comparison of three mortality prediction scores in emergency department patients with community-acquired pneumonia

Carolina Hincapié¹  Johana Ascuntar¹  Alba León¹  Fabián Jaimes^{1,2} 
fabian.jaimes@udea.edu.co

1 Universidad de Antioquia; GRAEPIC - Clinical Epidemiology Academic Group (Grupo Académico de Epidemiología Clínica); Medellín, Colombia. **2** Universidad de Antioquia, School of Medicine, Department of Internal Medicine, Medellín, Colombia



ACCESO ABIERTO

Citación: Hincapié C, Ascuntar J, León A, Jaimes F. **Neumonía adquirida en la comunidad: comparación de tres puntuaciones de predicción de mortalidad en el servicio de urgencias.** Colomb Méd (Cali), 2021; 52(4):e2044287
<http://doi.org/10.25100/cm.v52i4.4287>

Recibido: 14 Apr 2020
Revisado: 30 Mar 2021
Aceptado: 25 Aug 2021
Publicado: 23 Oct 2021

Palabras clave:
Sepsis, neumonía, reglas de decisión clínica, mortalidad

Keywords:
Sepsis, pneumonia, clinical decision rules, mortality.

Copyright: © 2021 Universidad del Valle



Resumen

Introducción:

El qSOFA es un nuevo puntaje propuesto para ayudar a identificar pacientes con sospecha de infección y con alta probabilidad de desarrollar complicaciones graves. Los criterios del qSOFA son similares a los evaluados en los puntajes de pronóstico usados tradicionalmente en neumonía (CRB-65 y CURB-65), pero no está claro cuál es mejor para predecir la mortalidad y la admisión a la UCI en pacientes con neumonía en el servicio de urgencias

Objetivo:

Comparar tres puntajes (CURB-65, CRB-65 y qSOFA) para determinar la mejor herramienta para identificar en servicios de urgencias a los pacientes con neumonía con mayor riesgo de mortalidad o ingreso en la unidad de cuidados intensivos (UCI).

Métodos:

Análisis secundario de datos de tres estudios de cohorte prospectivos con pacientes atendidos por urgencias con diagnóstico de neumonía en 5 hospitales de Colombia. Se realizó validación y comparación de la exactitud de los puntajes por medio de medidas de discriminación y de calibración.

Resultados:

Las cohortes 1, 2 y 3 incluyeron 158, 745 y 207 pacientes, con mortalidad de 32.3%, 17.2% y 18.4%, respectivamente. Se requirió la admisión a la UCI para 52.5%, 43.5% y 25.6% pacientes³, respectivamente. La mejor AUC-ROC para mortalidad fue para CURB-65 en la cohorte 3 (AUC-ROC= 0.67). La calibración de los modelos fue adecuada para los tres puntajes ($p > 0.05$).

Conflicto de interés:

Los autores declaran que no tienen ningún conflicto de interés por financiación o una relación personal que pudiera tener una influencia en lo reportado en este artículo.

Agradecimientos:

Los autores agradecemos al Dr. Juan Jose Hurtado, Dr. Hernan Aguirre, y la Dra. Ines Pachón por el soporte a este proyecto.

Financiación:

Proyecto financiado por COLCIENCIAS -Departamento Administrativo de Ciencia, Tecnología e Innovación (codigo 111556933362, 111556933448, 111556933334) y la Universidad de Antioquia

Autor de correspondencia:

Fabián Jaimes. E-mail address: fabian.jaimes@udea.edu.co Telephone: [57+4] 219 2432. Fax: [57+4] 263 8282. Direction postal: 050010E

Conclusiones:

Ninguno de estos puntajes demostró ser un predictor adecuado de mortalidad e ingreso en UCI. Además, el CRB 65 mostró la capacidad discriminativa más baja.

Abstract**Background:**

qSOFA is a score to identify patients with suspected infection and risk of complications. Its criteria are like those evaluated in prognostic scores for pneumonia (CRB-65 - CURB-65), but it is not clear which is best for predicting mortality and admission to the ICU.

Objective:

Compare three scores (CURB-65, CRB-65 and qSOFA) to determine the best tool to identify emergency department patients with pneumonia at increased risk of mortality or intensive care unit (ICU) admission.

Methods:

Secondary analysis of three prospective cohorts of patients hospitalized with diagnosis of pneumonia in five Colombian hospitals. Validation and comparison of the score's accuracies were performed by means of discrimination and calibration measures.

Results:

Cohorts 1, 2 and 3 included 158, 745 and 207 patients, with mortality rates of 32.3%, 17.2% and 18.4%, and admission to ICU was required for 52.5%, 43.5% and 25.6%, respectively. The best AUC-ROC for mortality was for CURB-65 in cohort 3 (AUC-ROC=0.67). The calibration was adequate ($p>0.05$) for the three scores.

Conclusions:

None of these scores proved to be an appropriate predictor for mortality and admission to the ICU. Furthermore, the CRB 65 exhibited the lowest discriminative ability.

Contribución del estudio

1) ¿Por qué se realizó este estudio?

El CURB-65, CRB-65 y qSOFA se diseñaron para identificar a los pacientes con mayor riesgo de complicaciones y mortalidad. Estas tres puntuaciones comparten variables clínicas en su composición y la neumonía extrahospitalaria es la principal causa de sepsis; por lo tanto, explorar las posibles diferencias en su desempeño como modelos de pronóstico tendría implicaciones para la práctica clínica.

2) ¿Cuáles fueron los resultados más relevantes del estudio?

No encontramos que la qSOFA, CURB-65 o CRB-65 fueran herramientas adecuadas para discriminar la mortalidad hospitalaria o el ingreso en UCI en tres cohortes de pacientes con neumonía extrahospitalaria, que ingresaron en los servicios de urgencias de 5 hospitales de referencia de Medellín, Colombia.

3) ¿Qué aportan estos resultados?

La qSOFA, CURB-65 y CRB-65 resultaron ser herramientas predictivas ineficaces de mortalidad e ingreso en UCI en nuestras cohortes, por lo que es necesario desarrollar y validar modelos predictivos de pronóstico de neumonía adquirida en la comunidad que sean útiles para la población colombiana.

Introducción

La neumonía tiene un lugar preponderante dentro de las causas más importantes de sepsis en el mundo, ocupando el segundo lugar de frecuencia en Colombia^{1,2}. Es sabido que la neumonía confiere un alto riesgo de mortalidad a nivel mundial^{3,4}. El análisis de situación de salud en Colombia reportó que entre 2005 y 2012, dentro de las causas de muerte por enfermedades transmisibles, la infección respiratoria aguda se encontraba en primer lugar con el 48.6%, representando 56.2% de las muertes por enfermedades transmisibles en mujeres y 43.1% en hombres⁵.

Por otra parte, la atención en salud de los pacientes con infecciones graves acarrea grandes costos para un estado y su sistema de salud y representa un gran reto para el enfoque clínico, al no disponer de un marcador pronóstico sencillo y específico que permita identificar tempranamente individuos en riesgo que ameritan una atención diferencial. De esto deriva la importancia de contar con herramientas útiles desde el punto de vista clínico para estimar el riesgo de muerte o de complicaciones en pacientes hospitalizados en el servicio de urgencias con sospecha de infección. Recientemente se propuso el puntaje qSOFA con el fin de identificar de manera temprana aquellos pacientes con peor pronóstico en sospecha de infecciones bacterianas agudas. Del mismo modo, se han realizado varios estudios con el fin de definir un puntaje predictor de mortalidad específicamente en neumonía y se ha utilizado de forma amplia el puntaje CURB-65 por su fácil aplicación, comparado con otros puntajes como el PSI (Pneumonia Severity Index)⁶. Recientemente, el tercer consenso en sepsis (SEPSIS 3) alentó la implementación de la puntuación qSOFA (evaluación rápida de falla orgánica relacionada con la sepsis) en pacientes adultos con sospecha de tener una infección bacteriana aguda para la identificación temprana de aquellos con peor pronóstico⁷. El Ministerio de Salud de Colombia⁸, así como la Sociedad Argentina de Enfermedades Infecciosas⁹ y el Instituto Mexicano del Seguro Social¹⁰ en sus guías para el manejo de pacientes con neumonía adquirida en la comunidad, recomiendan implementar CURB-65, a pesar de la falta de estudios locales para confirmar y validar esta recomendación⁷.

Dado que el CURB-65, el CRB-65 y el qSOFA se diseñaron con el fin de identificar aquellos pacientes con mayor riesgo de complicaciones y mortalidad, y comparten no sólo algunas variables clínicas en su composición sino también la población de estudio de NAC como principal causante de la sepsis, tendría implicaciones para la práctica clínica explorar potenciales diferencias en su desempeño como modelos predictivos de pronóstico. Así mismo, es necesario validar en poblaciones independientes cualquier modelo multivariable que se haya desarrollado con propósitos de pronóstico o diagnóstico para un problema clínico ¹¹. Por esto, se plantea en esta investigación la validación y comparación de los tres puntajes para determinar la mejor herramienta para identificar a los pacientes con neumonía en el departamento de emergencias que tienen un mayor riesgo de mortalidad o ingreso en la unidad de cuidados intensivos (UCI).

Materiales y Métodos

Diseño del estudio y escenario

Se realizó un análisis de tres estudios de cohorte prospectivos desarrollados entre 2013 y 2016 en cinco servicios de urgencias de la ciudad de Medellín: Hospital Universitario San Vicente Fundación (560 camas de hospitalización de adultos y 45 camas de UCI en 4 unidades), la Institución Prestadora de Servicios de Salud Universitaria IPS Universitaria Clínica León XIII (450 camas de hospitalización de adultos y 24 camas de UCI en 2 unidades), Hospital Pablo Tobón Uribe (360 camas de hospitalización de adultos y 40 camas de UCI en tres unidades), Hospital General de Medellín (442 camas) y Clínica Las Américas (304 camas) ¹²⁻¹⁴. La primera cohorte se reclutó del servicio de urgencias de tres hospitales de tercer nivel: Hospital Pablo Tobón Uribe, Hospital General de Medellín y Clínica Las Américas (2013-2016). El segundo fue de los servicios de urgencias de tres hospitales de tercer nivel, Hospital Universitario San Vicente Fundación, Institución Universitaria Clínica León XIII y Hospital Pablo Tobón Uribe (2014-2016), y el tercero fue del servicio de urgencias del Hospital Universitario San Vicente Fundación (2014-2016).

Fuente de los datos

Para cada una de las cohortes originales, auxiliares de investigación entrenados recolectaron los datos basados en los registros de historia clínica electrónica de forma sistemática, revisando todos los ingresos a las instituciones hospitalarias, tamizando a todos los pacientes que ingresaban por urgencias con diagnóstico de infección, sepsis, sepsis grave o choque. La definición de la fuente de la infección, la presencia de disfunción orgánica o de choque séptico se verificó con los datos extraídos de los registros de historia clínica en las primeras 6 horas. Para testificar la veracidad de los datos, la información era evaluada de forma periódica por los coinvestigadores. Esta información era consignada en formularios diseñados específicamente para cada una de las investigaciones y luego almacenada en bases de datos electrónicas. Dado que las cohortes eran prospectivas, la evaluación de los predictores fue independiente del conocimiento de los resultados de interés.

Adicionalmente fue necesario para esta investigación, recuperar el registro del BUN al ingreso hospitalario para pacientes del Hospital Universitario San Vicente Fundación y Hospital Pablo Tobón Uribe. Para la recolección de los datos se contó con la aprobación de los comités de ética de cada una de las instituciones participantes.

Participantes

Para el actual estudio se tomó como criterio de inclusión aquellos pacientes que hayan ingresado a los estudios con diagnóstico de neumonía. Para la cohorte 1, se utilizaron los criterios de infección de los Centros para el Control y la Prevención de Enfermedades (CDC) para la inclusión, la cohorte 2 consistió en sospecha de infección con al menos un criterio de disfunción orgánica y la cohorte 3 consistió en sospecha clínica de infección. Criterio

de exclusión común para las 3 cohortes: pacientes que fueron dados de alta temprana o referenciados a otra institución, y pacientes con órdenes de no reanimación o enfermedades terminales (Anexo 1). En el presente estudio no se utilizaron criterios de exclusión adicionales que no se hubieran considerado en los estudios originales¹²⁻¹⁴.

Variables

La mortalidad hospitalaria fue el desenlace primario y como resultado secundario se incluyó la admisión a UCI.

Predictores

qSOFA: este puntaje de predicción de gravedad incluye las variables de puntaje en la escala de coma de Glasgow ≤ 14 , presión arterial sistólica ≤ 100 mm Hg y frecuencia respiratoria ≥ 22 respiraciones por minuto; con un punto por cada variable, para una puntuación entre 0 y 3. Se plantea que la presencia de dos de esos tres criterios podría predecir la mortalidad en pacientes con sospecha de infección por fuera de la UCI⁷.

CURB-65: este puntaje incluye las variables confusión, urea > 7 mmol/L, frecuencia respiratoria ≥ 30 , presión sistólica < 90 o diastólica < 60 mmHg, edad ≥ 65 años; con un punto por cada variable, para una puntuación entre 0 y 5. Se plantea que la presencia de tres o más puntos podría predecir la mortalidad en pacientes neumonía adquirida en la comunidad¹⁵.

CRB-65: este puntaje incluye las variables confusión descritas anteriormente en el CURB-65 a excepción de la urea; con un punto por cada variable, para una puntuación entre 0 y 4. Se plantea que la presencia de tres o más puntos podría predecir la mortalidad en pacientes neumonía adquirida en la comunidad¹⁵.

Tamaño de muestra

Dado que es un análisis secundario de datos, no se realizó un cálculo de tamaño de muestra porque se trabajará con el conjunto de pacientes de las respectivas cohortes que cumplan con los criterios de inclusión. No obstante, se calcula el poder para la diferencia esperada en las áreas bajo la curva a partir de un número de pacientes fijo y considerando una probabilidad de error del tipo I (alfa) también fija en 0.05. El cálculo se basa en la fórmula descrita por Hanley y McNeil^{16,17}.

Con un tamaño de muestras fijas de 158, 745 y 207 pacientes, para las cohortes 1, 2 y 3 respectivamente, un alfa de 0.05 y tomando los valores observados de las AUC ROC como θ_1 : 0.7 y θ_2 : 0.77 (basados en lo encontrado en el estudio de Kolditz y colaboradores¹⁸ dado que carecemos de esta información en nuestro medio) encontramos una estimación de poder de 0.52, 0.98 y 0.62 respectivamente.

Métodos estadísticos

Las variables cuantitativas con distribución normal se presentan en medias y desviaciones estándar, mientras las que no presentaron distribución normal se expresan en medianas y rangos intercuartiles (RIQ).

Se realizó una validación y comparación de los tres modelos predictivos (CURB-65, CRB-65 y qSOFA) en términos de pronóstico. Para determinar la exactitud en la predicción de los modelos, es necesario examinar tanto la calibración, como la discriminación. La primera compara y establece la concordancia entre los eventos observados y los previstos, mientras que la discriminación establece la capacidad con la cual el puntaje distingue entre los individuos que experimentan o no el evento de interés¹⁹. Se determinó el desempeño de los puntajes en cuanto a discriminación con base en el área bajo la curva de características operativas del receptor (AUC-ROC). Se probó las diferencias entre AUC-ROC por medio del estadístico de DeLong-DeLong²⁰.

La calibración se determinó a través del grado de correspondencia dado por la prueba de bondad de ajuste de Hosmer-Lemeshow ($p > 0.05$). Adicionalmente, se realizaron curvas de calibración basados en los resultados de los modelos en cada una de las cohortes.

Luego se estimaron las características operativas para predicción de mortalidad y necesidad de UCI para cada uno de los puntajes, tomando como punto de corte 2 o más puntos para el qSOFA y 3 ó más tanto para el CURB-65 como para el CRB-65, basados en la propuesta original de los modelos que indicaban estos puntos de corte como alto riesgo de mortalidad. Así mismo, se analizó el desempeño de cada uno de los modelos predictores según los todos los posibles puntos de corte y se comparó con los puntos de cohorte originalmente propuestos. Para calcular la sensibilidad, la especificidad, los valores predictivos y las razones de probabilidad de los puntajes mencionados con sus respectivos puntos de cortes, se utilizó el teorema de Bayes asumiendo como prueba de referencia o estándar de oro los desenlaces de mortalidad y necesidad de UCI. En el análisis principal, los datos faltantes se consideraron valores anormales (escenario del peor de los casos). Adicionalmente, se realizó un análisis de sensibilidad con dos modelos adicionales: el mejor escenario, considerando los datos faltantes como valores normales, y con una regresión normal multivariante (MVN), técnica de imputación múltiple, tomando el BUN, edad, sexo, índice de Charlson, SOFA y fisiología aguda, edad, evaluación crónica de la salud II (APACHEII) como valores independientes. Los análisis estadísticos se realizaron con el software Stata 14[®]. Los resultados se presentan con sus respectivos intervalos de confianza (IC) del 95%, y se aplicó un nivel de significancia de $p < 0.05$. Se siguieron los estándares de publicación dados por el TRIPOD (*Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis*)¹¹.

Resultados

Se analizaron 158, 745 y 207 pacientes para las cohortes 1, 2 y 3 respectivamente. En cada una de las cohortes se encontraron medianas de edad de 70 (RIQ= 56-81), 66 (RIQ= 54-77) y 60 (RIQ= 44-75); el 58.9, 52.8 y 43% tenían 65 o más años, el 34.2, 48.9 y 44.4% eran de sexo femenino, el 52.5, 43.5 y 25.6% requirió ingreso a UCI, y el 32.3, 17.2 y 18.4% fallecieron

Tabla 1. Características generales y evolución de los participantes del estudio

	Cohorte 1 (n=158)	Cohorte 2 (n=745)	Cohorte 3 (n=207)
Características			
Mediana de edad (RIQ)	70 (56-81)	66 (54-77)	60 (44-75)
Género femenino	54 (34.2%)	364 (48.9%)	92 (44.4%)
Criterios CDC	158 (100%)	628 (84.75%)	187 (90.34%)
Gravedad			
Charlson (RIQ)	1 (0-3)	1 (0-2)	1 (0-2)
SOFA (RIQ)	4 (3-6)	4 (3-6)	3 (2-5)
APACHE II (RIQ)	17 (12-21)	15 (11-19)	13 (9-17)
Variables			
FR (RIQ)	24 (20- 28)	22 (19-27)	23 (19- 26)
PAS (RIQ)	110 (90-130)	113 (92-132)	120 (100-140)
PAD (RIQ)	60 (49-72)	68 (55-80)	76 (60-84)
PAM (RIQ)	76 (64-90)	83 (68-97)	91 (73-101)
Glasgow (RIQ)	15 (14-15)	15 (15-15)	15 (15-15)
BUN (RIQ)	n= 143 27.1 (16.3- 45.2)	n= 704 21.5 (14-33.4)	n= 206 18 (13-29)
≥65 años	93 (58.9%)	393 (52.8%)	89 (43%)
Puntajes			
qSOFA (RIQ)	1 (1-2)	1 (1-2)	1 (0-1)
CURB 65 (RIQ)	2 (2-3)	2 (1-3)	1 (1-2)
CRB 65 (RIQ)	2 (1-2)	1 (1-2)	1 (0-2)
Desenlaces			
UCI	83 (52.5%)	324 (43.5%)	53 (25.6%)
Muerte	51 (32.3%)	128 (17.2%)	38 (18.4%)

Abreviaturas: SOFA, Sequential Organ Failure Assessment; APACHE II Acute Physiology and Chronic Health Evaluation II; FR: frecuencia respiratoria; PAS, presión arterial sistólica; PAD, presión arterial diastólica; PAM, presión arterial media; BUN, nitrógeno ureico en sangre. Las variables cuantitativas se expresaron como medianas y su respectivo rango intercuartílico; las variables categóricas se muestran en frecuencias absolutas y relativas.

Tabla 2. Características microbiológicas

	Cohorte 1 (n=158)	Cohorte 2 (n=745)	Cohorte 3 (n=207)
Características			
Hemocultivo solicitado	151 (95.6%)	632 (84.8%)	175 (84.5%)
Hemocultivo positivo	35 (23.2%)	68 (10.8%)	16 (9.1%)
Microrganismo			
<i>Streptococcus pneumoniae</i>	13 (8.6%)	23 (3.6%)	6 (3.4%)
<i>Klebsiella pneumoniae</i>	6 (4%)	6 (1%)	1 (1%)
<i>Haemophilus influenzae</i>	2 (1.3%)	6 (1%)	2 (1.1%)
<i>Staphylococcus aureus</i>	4 (2.7%)	8 (1%)	2 (1%)
<i>Escherichia coli</i>	5 (3.3%)	9 (1.4%)	1 (1%)
<i>Pseudomonas aeruginosa</i>	2 (1.3%)	3 (1%)	0

durante la hospitalización, respectivamente (Tabla 1). De los pacientes de cada una de las cohortes, al 95.6, 84.8 y 84.5%, respectivamente, se les solicitó hemocultivo; de estos, al 23.2, 10.8 y 9.1% se les aisló germen. Los microorganismos más frecuentemente encontrados en cada una de las cohortes fueron: *Streptococcus pneumoniae*, *Klebsiella pneumoniae*, *Haemophilus influenzae* y *Escherichia coli* (Tabla 2).

Para el desenlace de ingreso a UCI, la discriminación fue mala para los tres puntajes en las tres cohortes. A partir del estadístico de DeLong-DeLong, se encontró diferencia estadísticamente significativa entre las AUC-ROC en las cohortes 1 y 2 ($p < 0.05$) (Figura 1), con un AUC-ROC de 0.59 para el qSOFA, 0.43 para el CURB-65 y 0.44 para el CRB-65 en la cohorte 1. Para el desenlace de mortalidad, la discriminación no fue adecuada en ninguno de los tres puntajes, en ninguna de las tres cohortes. A partir del estadístico de DeLong-DeLong, se encontró diferencia estadísticamente significativa en las AUC-ROC de las cohortes 1 y 2 (Figura 2), con

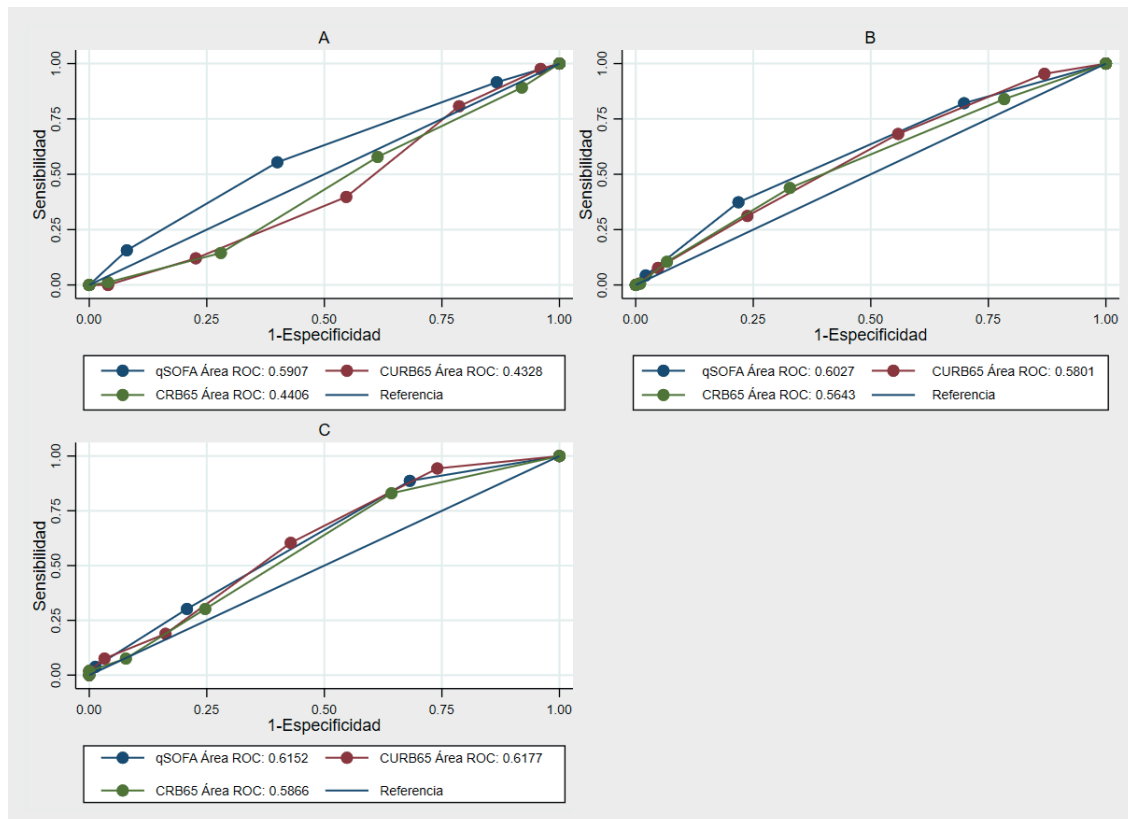


Figura 1. Curvas de las características operativas del receptor en las distintas cohortes para qSOFA, CURB-65 y CRB-65 en la discriminación de ingreso a UCI. A. Cohorte 1- DeLong-DeLong $p = 0.0008$. B. Cohorte 2- DeLong-DeLong $P = 0.0402$. C. Cohorte 3 - DeLong-DeLong $p = 0.3403$

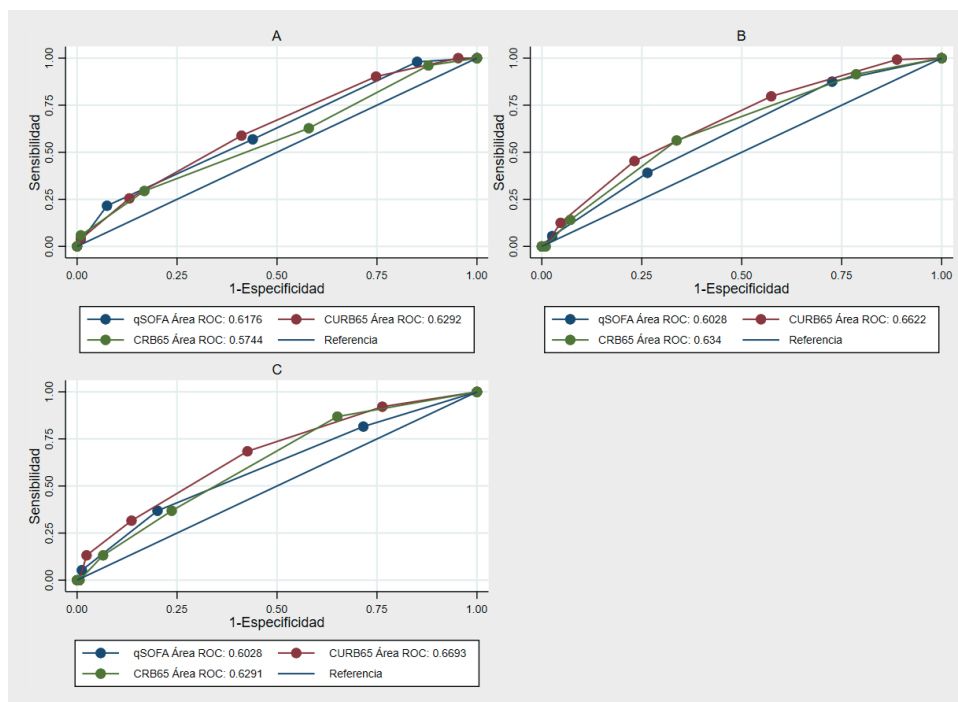


Figura 2. Curvas de las características operativas del receptor en las distintas cohortes para qSOFA, CURB-65 y CRB-65 en la discriminación de mortalidad. A. Cohorte 1 (DeLong-DeLong $p=0.0108$). B. Cohorte 2 (DeLong-DeLong $p=0.0218$). C. Cohorte 3 (DeLong-DeLong $p=0.1606$).

un AUC-ROC de 0.66 (IC del 95%= 0.62-0.71) para el CURB-65, 0.60 (IC del 95% = 0.56-0.65) para el qSOFA y 0.63 (IC del 95% = 0.59-0.68) para el CRB-65 en la cohorte 2.

La calibración de los modelos fue adecuada en la población de estudio para el ingreso en UCI y el resultado de mortalidad, según el estadístico de Hosmer-Lemeshow de los tres puntajes en cada una de las cohortes ($p > 0.05$) (Tabla S1).

Adicionalmente, se realizaron curvas de calibración para ambos resultados en los diferentes modelos en cada una de las cohortes, y se mostró un alto grado de correspondencia de las puntuaciones en la mayoría de las cohortes (Figura complementaria S1 y yS2S2).

En cuanto al desempeño de los modelos en sus características operativas, la mayor sensibilidad para la necesidad de UCI fue con el qSOFA (55.4%) y para la mortalidad fue con CURB-65 (58.8%) en la cohorte 1. La mayor especificidad fue con CRB-65 para tanto la necesidad de UCI como la mortalidad, con 93.5% y 93.4% en las cohortes 2 y 3, respectivamente. El menor desempeño en la predicción de mortalidad en términos de sensibilidad fue para el CRB-65 en la cohorte 3 (13.2%), para la especificidad fue para el qSOFA en la cohorte 1 (43.9%) y para el valor predictivo positivo fue el CRB-65. en la cohorte 3 (Tablas S2 y yS3S3).

Discusión

Encontramos que qSOFA, CURB-65 o CRB-65 no fueron óptimos para discriminar la mortalidad hospitalaria o el ingreso a la UCI en tres cohortes de pacientes con neumonía adquirida en la comunidad ingresados en cinco hospitales de Medellín. Sin embargo, al observar el AUC, la sensibilidad y los valores de valor predictivo negativo, CURB-65 pareció funcionar consistentemente mejor que las otras dos herramientas con respecto a la discriminación de la mortalidad. Por el contrario, con respecto a la calibración, fue posible demostrar un buen desempeño para las tres puntuaciones en las 3 cohortes. No obstante, la falta de un buen desempeño discriminativo indica que estos sistemas de puntuación no deben utilizarse como herramientas predictivas^{19,21}.

Es necesario tener en cuenta el entorno de los estudios que originalmente desarrollaron las puntuaciones: el CURB-65 y el CRB-65 se desarrollaron en Reino Unido, Nueva Zelanda y Holanda ²² hace más de 20 años, países con una mortalidad asociada a neumonía adquirida en la comunidad menor que en Colombia (9% vs 17-32%). Por otro lado, el qSOFA se derivó de una cohorte muy reciente ²³ que utilizó un espectro clínico más allá de la neumonía y mostró una mortalidad hospitalaria de solo el 4%. En 2006, Capelastegui *et al.* ²⁴, mostró un desempeño similar entre las puntuaciones CURB-65 y CRB-65 para la mortalidad a los 30 días con un AUC superior a 0,85. Posteriormente, Man *et al.* compararon estas reglas de predicción para la mortalidad a 30 días en pacientes con neumonía adquirida en la comunidad y encontraron AUC más altas que las observadas en el presente estudio ²⁵. En los estudios originales que sirvieron de base para el desarrollo de qSOFA, Seymour *et al.* encontraron un buen desempeño para la predicción de la mortalidad intrahospitalaria ²⁶. Posteriormente, Wang *et al.* realizaron un análisis secundario de datos de una cohorte prospectiva donde evaluaron el desempeño de qSOFA en pacientes con diagnóstico de infección que ingresaron al servicio de urgencias, y encontraron que la puntuación no tuvo un buen desempeño (AUC= 0.66) para la mortalidad a 28 días ²⁷.

Estudios anteriores han demostrado que estas puntuaciones subestiman el riesgo en pacientes con neumonía adquirida en la comunidad. Hace un par de años, Chen *et al.* compararon el desempeño de la qSOFA, CRB y CRB-65 con respecto a la mortalidad y el ingreso en UCI. Los valores de AUC-ROC del qSOFA para la predicción de la mortalidad a los 28 días fueron similares a los de las puntuaciones y CRB-65, 0.655 vs 0,661 respectivamente. Asimismo, la predicción de ingreso en UCI mostró medidas de discriminación similares, 0.666 vs 0.685 respectivamente ²⁸. Estos resultados son consistentes con los nuestros en cuanto a discriminación, tanto por ingreso en UCI como por mortalidad, a pesar de su gran tamaño muestral y de ser realizado en un solo hospital, lo que podría resultar en una menor variabilidad en la muestra global. En Alemania, Kolditz *et al.* ¹⁸, compararon qSOFA con CRB y CRB-65 para la mortalidad a 30 días en pacientes con neumonía extrahospitalaria. Encontraron que los resultados de AUC-ROC favorecieron a CRB-65 en comparación con qSOFA, AUC-ROC 0,77 frente a 0.70 respectivamente.

Más recientemente, tres estudios diferentes ²⁹⁻³¹ muestran la comparación de las puntuaciones de gravedad en pacientes con neumonía COVID-19, todos ellos muestran que CURB-65 podría ser mejor comparado con qSOFA para estimar la mortalidad. Guohui y col. ³⁰, encontraron un AUC para la mortalidad por descarga de 0.85 para CURB-65, 0.80 para CRB-65 y 0.73 para qSOFA. Bradley *et al.*, encontraron un AUC para 30 días de mortalidad de 0.75 para CURB-65 y 0.62 para qSOFA. Lazar *et al.*, encontraron un AUC durante 30 días en la mortalidad hospitalaria de 0.74 para CURB-65 y 0.63 para qSOFA.

Como se muestra en los estudios presentados anteriormente, el desempeño de los puntajes varió significativamente entre todas las cohortes debido a sus diferencias, incluyen la distribución de agentes etiológicos, enfermedades coexistentes, apoyo social, disponibilidad de recursos y conductas médicas, incluyendo los criterios de ingreso a UCI. En nuestro estudio, el desempeño de estas puntuaciones varió a pesar de que las cohortes eran de la misma ciudad, lo que puede explicarse por la variabilidad en los criterios de inclusión de pacientes.

El AUC-ROC es un parámetro estadístico que permite la comparación de modelos predictivos de diagnóstico o pronóstico en términos de capacidad de discriminación, y es razonable utilizar un AUC-ROC >0.75 como referencia de rendimiento aceptable. Sin embargo, esta medida estadística no permite una interpretación clínica directa, y esta limitación en los modelos predictivos es una constante en la literatura sobre este tema, por lo que siempre es necesario evaluar simultáneamente sus características operativas. En cuanto a la calibración, ninguno de los estudios mencionados anteriormente tuvo en cuenta esto en el análisis estadístico. A menudo se subestima la importancia fundamental de una mala calibración.

Esto puede conducir a una disminución de la utilidad clínica; la implementación de una herramienta predictiva con mala calibración podría incluso llevar a la toma de decisiones perjudiciales para el paciente³². Los estudios futuros podrían considerar otras variables para el cálculo de la puntuación, como variables relacionadas con el agente microbiológico, oximetría de pulso, temperatura y comorbilidades como enfermedad pulmonar obstructiva crónica, insuficiencia cardíaca congestiva e inmunosupresión, entre otras.

Una de las limitaciones de nuestro estudio fue el tamaño de la muestra. Basamos la diferencia de 0.7-0.77 entre la discriminación (AUC-ROC) de CRB-65 y las puntuaciones de qSOFA en información parcial de Kolditz *et al*¹⁸. Sin embargo, esta diferencia no tiene necesariamente una base clínica y no consideró que todas las puntuaciones tuvieran un rendimiento final de discriminación pobre (AUC <0.75). La aproximación tradicional del cálculo del tamaño de la muestra en modelos predictivos define un valor de al menos 10 resultados para cada variable independiente^{33,34}. Para las comparaciones entre modelos, exclusivamente mediante discriminación, basamos la fórmula del tamaño de la muestra en la comparación AUC-ROC de Hanley-McNeil^{16,17}. Sin embargo, específicamente para la validación de modelos predictivos, no existe una indicación clara del cálculo del tamaño de la muestra, y aunque algunos autores han sugerido un mínimo de 100 resultados, muchos estudios no consideran este aspecto^{35,36}. Por otro lado, la recolección se realizó en cinco instituciones que son reconocidas como centros de atención de salud de alta calidad, lo que puede llevar a un sesgo de selección. Sin embargo, las tres cohortes tenían diferentes criterios de inclusión, lo que mejoró significativamente el espectro clínico de la población de estudio. Otra limitación fue que, a pesar de ser cohortes construidas prospectivamente, este estudio proporciona un análisis secundario de los datos, lo que da lugar a valores de urea faltantes para algunos participantes. Estos datos faltantes se consideraron valores anormales, lo que podría generar un sesgo de clasificación diferencial o no diferencial. Sin embargo, los datos faltantes representaron solo el 5% y el análisis de sensibilidad con diferentes escenarios no mejoró el desempeño de los modelos.

Un modelo predictivo no es de uso práctico si no puede discriminar y calibrar al mismo tiempo: separar adecuadamente a los que presentan la condición de los que no, es tan importante como si existe concordancia entre los eventos observados y esperados²¹. A diferencia de la supervisión requerida para las nuevas tecnologías médicas, los sistemas de predicción no están sujetos a juicios estrictos, a pesar del riesgo potencial de afectar a un mayor número de pacientes debido a su extensa implementación.

Conclusión

En las tres cohortes independientes de pacientes ingresados en el servicio de urgencias con neumonía, se encontró que qSOFA, CURB-65 y CRB-65 eran herramientas de predicción limitadas para la mortalidad y el ingreso en la UCI. Además, el CRB-65 exhibió la capacidad discriminativa más baja.

Referencias

1. Rodríguez F, Barrera L, De La Rosa G, Dennis R, Duenas C, Granados M, et al. The epidemiology of sepsis in Colombia: a prospective multicenter cohort study in ten university hospitals. *Crit Care Med*. 2011;39(7):1675-82. Doi: 10.1097/CCM.0b013e318218a35e.
2. Angus DC, van der Poll T. Severe sepsis and septic shock. *New Engl J Med*. 2013; 369(9): 840-51. Doi: 10.1056/NEJMra1208623.
3. Lindenauer PK, Lagu T, Shieh MS, Pekow PS, Rothberg MB. Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003-2009. *JAMA*. 2012; 307(13): 1405-13. Doi: 10.1001/jama.2012.384.

4. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007; 44 Suppl 2: S27-72. Doi: 10.1086/511159.
5. Ministerio de Salud y Protección Social. Análisis de Situación de salud de Colombia 2014. Bogotá: Ministerio de Salud y Protección Social; 2014. Available in: https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/PSP/ASIS_2014_v11.pdf.
6. Chalmers JD, Singanayagam A, Akram AR, Mandal P, Short PM, Choudhury G, et al. Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia. Systematic review and meta-analysis. *Thorax*. 2010; 65(10): 878-83. Doi: 10.1136/thx.2009.133280.
7. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016; 315(8): 801-10. Doi: 10.1001/jama.2016.0287.
8. Ministerio de la Protección Social. Guías para Manejo de Urgencias. 3a Edición. Bogotá: Ministerio de la Protección; 2009.
9. Lopardo G, Basombrío A, Clara L, Desse J, De Vedia L, Libero E, et al. Neumonía adquirida de la comunidad en adultos: recomendaciones sobre su atención. *Medicina (Buenos Aires)*. 2015;75:245-57.
10. Instituto Mexicano del Seguro Social. Prevención, Diagnóstico y Tratamiento de la Neumonía Adquirida en la Comunidad. Evidencias y Recomendaciones. Catálogo Maestro de Guías de Práctica Clínica: IMSS-234-09. México, DF:
11. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015; 350: g7594. Doi: 10.1136/bmj.g7594.
12. Londono J, Nino C, Archila A, Valencia M, Cardenas D, Perdomo M, et al. Antibiotics has more impact on mortality than other early goal-directed therapy components in patients with sepsis: An instrumental variable analysis. *J Crit Care*. 2018;48:191-7. Doi: 10.1016/j.jcrc.2018.08.035.
13. Castano P, Plaza M, Molina F, Hincapie C, Maya W, Catano J, et al. Antimicrobial agent prescription: a prospective cohort study in patients with sepsis and septic shock. *Trop Med Internat Health*. 2019; 24(2): 175-84. Doi: 10.1111/tmi.13186.
14. Londono J, Nino C, Diaz J, Morales C, Leon J, Bernal E, et al. Association of clinical hypoperfusion variables with lactate clearance and hospital mortality. *Shock*. 2018; 50(3): 286-92. Doi: 10.1097/SHK.0000000000001066.
15. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;58(5):377. Doi: 10.1136/thorax.58.5.377.
16. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143(1): 29-36. Doi: 10.1148/radiology.143.1.7063747.
17. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*. 1983;148(3):839-43. Doi: 10.1148/radiology.148.3.6878708.
18. Kolditz M, Scherag A, Rohde G, Ewig S, Welte T, Pletz M. Comparison of the qSOFA and CRB-65 for risk prediction in patients with community-acquired pneumonia. *Intensive Care Med*. 2016;42(12):2108-10. Doi: 10.1007/s00134-016-4517-y.
19. Altman D, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453 - 73. Doi: 10.1002/(SICI)1097-0258(20000229)19:4 <453::AID-SIM350 >3.0.CO;2-5

20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44(3):837-45.
21. Cook NR. Statistical Evaluation of Prognostic versus Diagnostic Models: Beyond the ROC Curve. *Clin Chem*. 2008;54(1):17-23. Doi: 10.1373/clinchem.2007.096529
22. Lim WS, Macfarlane JT, Boswell TC, Harrison TG, Rose D, Leinonen M, et al. Study of community acquired pneumonia aetiology (SCAPA) in adults admitted to hospital: implications for management guidelines. *Thorax*. 2001;56(4):296-301. Doi: 10.1136/thorax.56.4.296.
23. Ranzani OT, Prina E, Menendez R, Ceccato A, Cilloniz C, Mendez R, et al. New Sepsis Definition (Sepsis-3) and Community-acquired Pneumonia Mortality. A Validation and Clinical Decision-Making Study. *Am J Respir Crit Care Med*. 2017; 196(10): 1287-97. Doi: 10.1164/rccm.201611-2262OC.
24. Capelastegui A, Espana PP, Quintana JM, Areitio I, Gorordo I, Egurrola M, et al. Validation of a predictive rule for the management of community-acquired pneumonia. *Europ Respirat J*. 2006;27(1):151-7. Doi: 10.1183/09031936.06.00062505.
25. Man SY, Lee N, Ip M, Antonio GE, Chau SS, Mak P, et al. Prospective comparison of three predictive rules for assessing severity of community-acquired pneumonia in Hong Kong. *Thorax*. 2007;62(4):348-53. Doi: 10.1136/thx.2006.069740
26. Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):762-74. Doi: 10.1001/jama.2016.0288.
27. Wang JY, Chen YX, Guo SB, Mei X, Yang P. Predictive performance of quick Sepsis-related Organ Failure Assessment for mortality and ICU admission in patients with infection at the ED. *Am J Emerg Med*. 2016;34(9):1788-93. Doi: 10.1016/j.ajem.2016.06.0152
28. Chen YX, Wang JY, Guo SB. Use of CRB-65 and quick Sepsis-related Organ Failure Assessment to predict site of care and mortality in pneumonia patients in the emergency department: a retrospective study. *Crit Care*. 2016;20(1):167. Doi: 10.1186/s13054-016-1351-0.
29. Lazar Neto F, Marino LO, Torres A, Cilloniz C, Meirelles Marchini JF, Garcia de Alencar JC, et al. Community-acquired pneumonia severity assessment tools in patients hospitalized with COVID-19: a validation and clinical applicability study. *Clin Microbiol Infect*. 2021; 27(7) :1037.e1-1037.e8. Doi: 10.1016/j.cmi.2021.03.002.
30. Fan G, Tu C, Zhou F, Liu Z, Wang Y, Song B, et al. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. *Eur Respir J*. 2020;56(3): 2002113. Doi: 10.1183/13993003.02113-2020
31. Bradley P, Frost F, Tharmaratnam K, Wootton DG, Research NWCOFR. Utility of established prognostic scores in COVID-19 hospital admissions: multicentre prospective evaluation of CURB-65, NEWS2 and qSOFA. *BMJ Open Respir Res*. 2020;7(1):e000729. Doi: 10.1136/bmjresp-2020-000729.
32. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. *Medical Decision Making*. 2015;35(2):162-9. Doi: 10.1177/0272989X14547233.
33. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*. 2007;165(6):710-8. Doi: 10.1093/aje/kwk052.
34. Courvoisier DS, Combescure C, Agoritsas T, Gayet-Ageron A, Perneger TV. Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure. *J Clin Epidemiol*. 2011;64(9):993-1000. Doi: 10.1016/j.jclinepi.2010.11.012.
35. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58(5):475-83. Doi: 10.1016/j.jclinepi.2004.06.017.
36. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567-86. Doi: 10.1002/sim.1844.

Anexo 1. Criterios de inclusión y exclusión.

Cohorte 1

Criterios de inclusión: pacientes con una infección de acuerdo con los criterios clínicos o microbiológicos de las definiciones de los CDC, y al menos una disfunción de un órgano / sistema, basada en el puntaje SOFA ≥ 2 , producida o relacionada con la infección, y detectada dentro de las 24 h previas a ser admitido en el estudio. Para la HGM y CLA, además de los criterios anteriores, los pacientes debían tener hemocultivos positivos. Esta condición se estableció debido a problemas en el reclutamiento y la necesidad de mejorar la eficiencia en el análisis por parte de los especialistas en enfermedades infecciosas.

Criterios de exclusión: Se excluyeron los pacientes remitidos a otro hospital en las primeras 48 h del ingreso, los pacientes con infecciones conocidas que necesitaban tratamientos prolongados (p. Ej., Tuberculosis, nocardiosis, histoplasmosis) y los pacientes con órdenes de no reanimación o enfermedades terminales.

Cohorte 2

Criterios de inclusión: Pacientes ≥ 18 años, ingresados en urgencias con diagnóstico sospechado o confirmado de infección, sepsis, sepsis grave o shock séptico; al menos dos criterios de síndrome de respuesta inflamatoria sistémica y presión arterial sistólica < 90 mmHg después de un bolo de cristaloides de al menos 20 ml/kg, o un lactato sérico > 4 mmol / L.

Criterios de exclusión: Negativa del paciente, la familia o el médico tratante a participar; embarazo, infarto de miocardio, accidente cerebrovascular, crisis asmática, arritmia, traumatismo, hemorragia gastrointestinal, convulsiones no debidas a meningitis, sobredosis de sustancias psicoactivas, cirugía < 24 horas, quemaduras, recuento de CD4 < 50 células / mm³, estado hiperosmolar, cetoacidosis diabética o cirrosis; dado de alta o referido en las primeras 24 horas, participación previa en el estudio, derivación de otra institución donde el paciente ha estado hospitalizado > 24 horas, o una orden de no reanimar.

Cohorte 3

Criterios de inclusión: pacientes ≥ 18 años ingresados en Urgencias con una infección bacteriana aguda confirmada por evidencia clínica o de laboratorio, de acuerdo con los criterios de los CDC; Además, los pacientes deben estar disponibles para evaluar los signos vitales y los hallazgos físicos al ingresar en el hospital.

Criterios de exclusión: pacientes remitidos desde otra institución donde permanecieron más de 24 horas; alta o referencia a otra institución durante las primeras 24 horas de ingreso; enfermedades que dificultan la evaluación física de parámetros clínicos, como amputación de miembros, quemaduras extensas o enfermedades cutáneas graves, fenómeno de Raynaud o enfermedad arterial periférica; cirrosis; trombosis mesentérica; la negativa del paciente a participar; detección después de 6 horas de haber sido admitido en la sala de emergencias; una participación previa en el estudio y una orden de no reanimación.

Tablas

Tabla S1. Calibración de los modelos para ingreso a UCI y mortalidad

	Cohorte 1 (n=158)	Cohorte 2 (n=745)	Cohorte 3 (n=207)
Ingreso UCI			
qSOFA			
Número de grupos	7	5	7
Hosmer-Lemeshow	1.6	1.1	2.0
Valor P	0.9002	0.7875	0.8449
CURB-65			
Número de grupos	10	10	9
Hosmer-Lemeshow	8	6.2	3.1
Valor P	0.4355	0.6284	0.8749
CRB-65			
Número de grupos	10	8	7
Hosmer-Lemeshow	12.1	3.9	3.3
Valor P	0.1466	0.6941	0.6478
Mortalidad			
qSOFA			
Número de grupos	6	6	4
Hosmer-Lemeshow	4.6	1.1	0.14
Valor P	0.3365	0.8985	0.9316
CURB-65			
Número de grupos	10	10	9
Hosmer-Lemeshow	5	8.1	3.5
Valor P	0.7612	0.4236	0.8309
CRB-65			
Número de grupos	10	7	7
Hosmer-Lemeshow	5.7	6.42	1.9
Valor p	0.6843	0.2672	0.8632

Tabla S2. Características operativas para predicción de necesidad de UCI de los puntajes qSOFA (con punto de corte ≥ 2), CURB 65 (con punto de corte ≥ 3) y CRB 65 (con punto de corte ≥ 3).

	Sensibilidad (95% IC)	Especificidad (95% IC)	VPP (95% IC)	VPN (95% IC)	LR + (95% IC)	LR - (95% IC)
qSOFA						
Cohorte 1	55.4 (44.1- 66.7)	60.0 (48.3-71.8)	60.5 (48.9- 72.2)	54.9 (43.5- 66.3)	1.4 (1-1.9)	0.7 (0.6- 1.0)
Cohorte 2	37.4 (31.9- 42.8)	78.2 (74.1-82.2)	56.8 (49.9- 63.7)	61.8 (57.6- 66.1)	1.7 (1.4- 2.2)	0.8 (0.7-0.9)
Cohorte 3	30.2 (16.9-43.5)	79.2 (72.5-86)	33.3 (19- 47.7)	76.7 (69.9- 83.6)	1.5 (0.9-2.4)	0.9 (0.7-1.1)
CURB-65						
Cohorte 1	39.8 (28.6-50.9)	45.3 (33.4-57.3)	44.6 (32.6- 56.6)	40.5 (29.4- 51.6)	0.7 (0.5-1.0)	1.3 (1-1.8)
Cohorte 2	31.2 (26- 36.4)	76.3 (72.1-80.4)	50.3 (43.1- 57.4)	59.0 (54.8-63.2)	1.3 (1.0-1.7)	0.9 (0.8- 1)
Cohorte 3	18.9 (7.4- 30.3)	83.8 (77.6- 89.9)	28.6 (12.2- 45)	75.0 (68.2- 81.8)	1.2 (0.6- 2.3)	1 (0.8- 1.1)
CRB-65						
Cohorte 1	14.5 (6.3- 22.6)	72.0 (61.2- 82.8)	36.4 (18.4- 54.3)	43.2 (34.1- 52.3)	0.5 (0.3- 1)	1.2 (1.0- 1.4)
Cohorte 2	10.5 (7.0- 14)	93.4 (90.9- 95.9)	54.9 (41.6- 68.0)	57.5 (53.8- 61.3)	1.6 (1- 2.6)	1 (0.9- 1.0)
Cohorte 3	7.6 (0- 15.6)	92.2 (87.7- 96.8)	25.0 (0.7- 49.3)	74.4 (67.9- 80.8)	1 (0.3- 2.9)	1.0 (0.9- 1.1)

Tabla S3. Características operativas para predicción de mortalidad de los puntajes qSOFA (con punto de corte ≥ 2), CURB 65 (con punto de corte ≥ 3) y CRB 65 (con punto de corte ≥ 3).

	Sensibilidad (95% IC)	Especificidad (95% IC)	VPP (95% IC)	VPN (95% IC)	LR + (95% IC)	LR - (95% IC)
qSOFA						
Cohorte 1	56.9 (42.3- 71.4)	43.9 (34.1- 53.8)	32.6 (22.3- 42.9)	68.1 (56.4- 79.8)	1.0 (0.8- 1.4)	1 (0.7- 1.4)
Cohorte 2	39.1 (30.2- 47.9)	73.6 (70.0- 77.1)	23.5 (17.6- 29.4)	85.3 (82.2- 88.4)	1.5 (1.2- 1.9)	0.8 (0.7- 1)
Cohorte 3	36.8 (20.2- 53.5)	79.9 (73.5- 86.2)	29.2 (15.3- 43.1)	84.9 (79.0- 90.8)	1.8 (1.1- 3.1)	0.8 (0.6- 1.0)
CURB-65						
Cohorte 1	58.8 (44.3-73.3)	58.9 (49.1-68.7)	40.5 (28.7-52.4)	75.0 (65.1-84.9)	1.4 (1.0-2)	0.7 (0.5-1.0)
Cohorte 2	45.3 (36.3-54.3)	76.8 (73.4- 80.2)	28.9 (22.3- 35.4)	87.1 (84.2- 90.0)	2 (1.5- 2.3)	0.7 (0.6- 0.8)
Cohorte 3	31.6 (15.5- 47.7)	86.4 (80.9- 91.9)	34.3 (17.1- 51.4)	84.9 (79.2- 90.5)	2.3 (1.3- 4.2)	0.8 (0.6- 1)
CRB-65						
Cohorte 1	29.4 (15.9-42.9)	83.2 (75.6-90.7)	45.5 (27-64)	71.2 (62.9-79.5)	1.8 (1-3.2)	0.9 (0.7-1.0)
Cohorte 2	14.1 (7.7-20.5)	92.9 (90.9-95)	29.0 (16.9-41.1)	83.9 (81.1-86.7)	2 (1.1-3.3)	0.9 (0.7-1)
Cohorte 3	13.2 (1.1-25.2)	93.5 (89.5-97.5)	31.3 (5.4-57.1)	82.7 (77.1-88.4)	2.0 (0.8-5.5)	0.9 (0.8-1.1)

Figuras

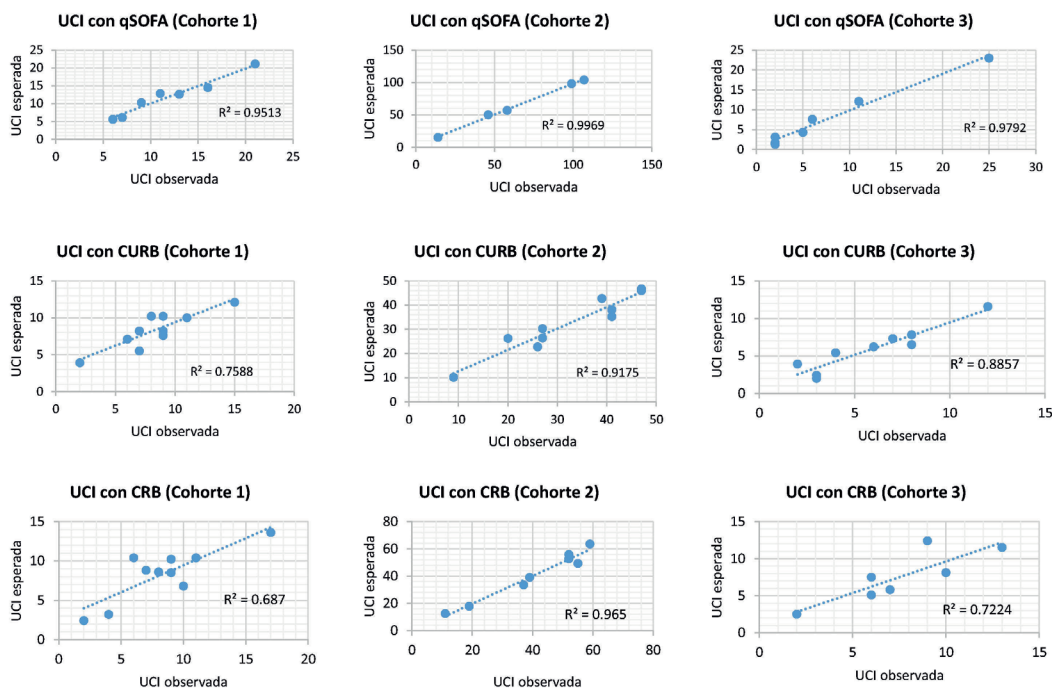


Figura S1. Curvas de calibración para para predicción de necesidad de UCI

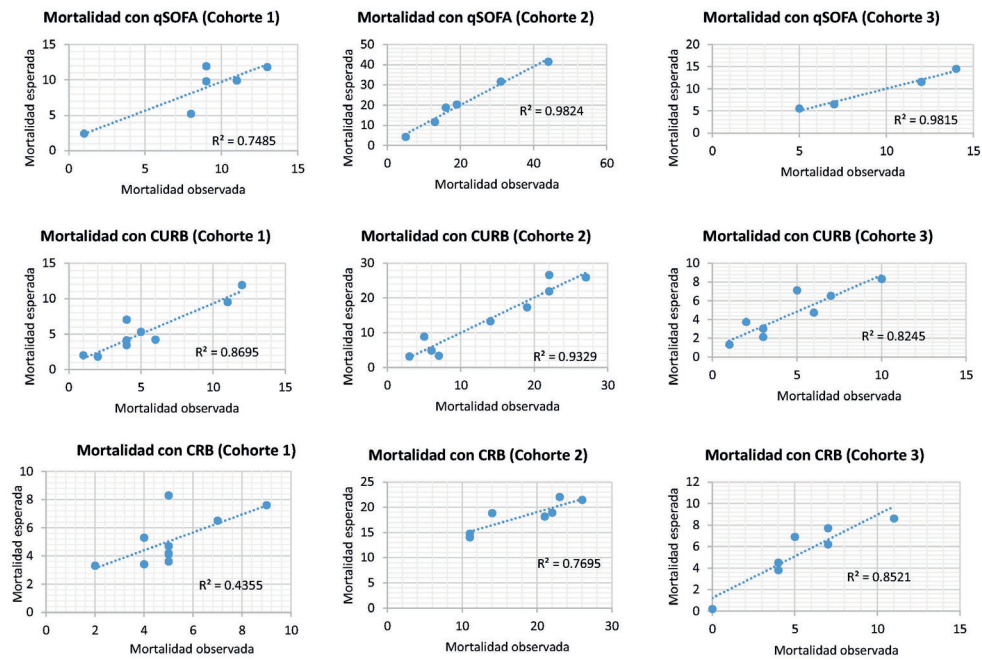


Figura S2. Curvas de calibración para para predicción de mortalidad