



ORIGINAL

Spanish Influenza Score (SIS): utilidad del *Machine Learning* en el desarrollo de una escala temprana de predicción de mortalidad en la gripe grave



Grupo de Trabajo Gripe A Grave (GETGAG) de la Sociedad Española de Medicina Intensiva Crítica y Unidades Coronarias (SEMICYUC) [◇]

Recibido el 16 de marzo de 2020; aceptado el 23 de mayo de 2020
 Disponible en Internet el 11 de agosto de 2020

PALABRAS CLAVE

Gripe grave;
 Pronóstico;
Machine learning

Resumen

Objetivo: Desarrollar una escala predictiva de mortalidad (SIS) en pacientes con gripe grave considerando las variables de ingreso a UCI y comparar su eficacia respecto de un modelo de Random Forest (RF).

Diseño: Sub-análisis de base de datos GETGAG/SEMICYUC.

Ámbito: Medicina intensiva.

Intervenciones: Ninguna.

Pacientes: Pacientes ingresados en 184 UCI españolas (2009-2018) con infección por gripe.

Variables: Demográficas, nivel de gravedad, tiempo síntomas hasta el ingreso al hospital (Gap-H) o desde hospital a UCI (Gap-UCI), o al diagnóstico (Gap-Dg), vacunación, cuadrantes infiltrados, insuficiencia renal, ventilación no-invasiva o invasiva (VM), shock y comorbilidades. Los puntos de corte y la importancia de las variables se obtuvieron de forma automática. Se realizó validación cruzada y regresión logística a partir de la cual se desarrolló la puntuación SIS. Se aplicó la puntuación y se calculó la exactitud y la discriminación (AUC-ROC) para SIS y RF. El análisis se realizó mediante CRAN-R Project.

Resultados: Se incluyeron 3.959 pacientes, edad 55 (43-67) años, 60% hombres, APACHE II de 16 (12-21) y SOFA 5 (4-8) puntos y una mortalidad del 21,3%. VM, shock, APACHE II, SOFA, insuficiencia renal aguda y Gap-UCI fueron incluidas en SIS. A partir de los OR se construyó el SIS que demostró una exactitud del 83% y un AUC-ROC del 82% similar al RF (AUC-ROC 82%).

Conclusiones: La escala SIS de fácil aplicación, ha demostrado una adecuada capacidad de estratificación del riesgo de mortalidad en la UCI. Sin embargo, estos resultados deberán ser validados prospectivamente.

© 2020 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.

[◇] En el [anexo](#) al final del artículo se relacionan los nombres de todos los autores.

KEYWORDS

Severe influenza;
Prognosis;
Machine learning

Spanish Influenza Score (SIS): Usefulness of machine learning in the development of an early mortality prediction score in severe influenza

Abstract

Objective: To develop a mortality prediction score (Spanish Influenza Score [SIS]) for patients with severe influenza considering only variables at ICU admission, and compare its performance respect of Random Forest (RF).

Design: Sub-analysis from the GETGAG/SEMICYUC database.

Scope: Intensive Care Medicine.

Patients: Patients admitted to 184 Spanish ICUs (2009-2018) with influenza infection Intervention: None.

Variables: Demographic data, severity of illness, times from symptoms onset until hospital admission (Gap-H), hospital to ICU (Gap-ICU) or hospital to diagnosis (Gap-Dg), antiviral vaccination, number of quadrants infiltrated, acute renal failure, invasive or noninvasive ventilation, shock and comorbidities. The study variable cut-off points and importance were obtained automatically. Logistic regression analysis with cross-validation was performed to develop the SIS score using the output coefficients. Accuracy and discrimination (AUC-ROC) were applied to evaluate SIS and RF. All analyses were performed using R (CRAN-R Project).

Results: A total of 3959 patients were included. The mean age was 55 years (range 43-67), 60% were men, APACHE II 16 (12-21) and SOFA 5 (4-8), with ICU mortality 21.3%. Mechanical ventilation, shock, APACHE II, SOFA, acute renal failure and Gap-ICU were included in the SIS. The latter was generated according to the ORs obtained by logistic regression, and showed an accuracy of 83% with an AUC-ROC of 82%, similar to RF (AUC-ROC 82%).

Conclusions: The SIS score is easy to apply and shows adequate capacity to stratify the risk of ICU mortality. However, further studies are needed to validate the tool prospectively.

© 2020 Elsevier España, S.L.U. y SEMICYUC. All rights reserved.

Introducción

La mortalidad de los pacientes críticos ingresado en UCI afectados por infección del virus de la gripe continúa siendo inaceptablemente elevada, algo superior al 20% en la población general¹⁻³ y mayor al 30% en aquellos pacientes que requieren ventilación mecánica invasiva⁴. Las escalas utilizadas para predecir la gravedad en pacientes con neumonía comunitaria parecen subestimar la gravedad de los pacientes con infección por gripe⁵. La posibilidad de contar con predictores precoces de evolución, puede ser una herramienta de ayuda a la decisión clínica, al momento de la asistencia de estos pacientes críticos. Diferentes estudios han intentado descubrir predictores relacionados con la mortalidad de este colectivo particular de pacientes en UCI⁵⁻⁸. Sin embargo, la mayoría de ellos presentan graves limitaciones relacionadas con el pequeño número de pacientes incluidos^{6,7}, la metodología empleada para obtener el predictor^{5,8} o aplicación a subpoblaciones especiales^{9,10}. Desarrollar predictores de mortalidad en pacientes críticos es una tarea compleja, debido a su heterogeneidad y a la diferente respuesta sistémica ante la misma enfermedad. Las nuevas tecnologías informáticas permiten generar de forma automática, modelos de predicción mediante la utilización de estrategias de «machine learning»¹¹⁻¹⁵. Sin embargo, la mayoría de estos modelos son poco comprensibles por los médicos, con una muy baja aceptación a las decisiones clínicas basadas en algoritmos crípticos (cajas negras) y en general, no tienen una clara aplicación en

la práctica clínica¹⁶. Por ello, es importante desarrollar modelos predictivos que aprovechen estas nuevas tecnologías de análisis pero que sean comprensibles, precoces y de aplicación práctica, con la finalidad de ayudar en la decisión clínica. El objetivo de nuestro estudio es desarrollar mediante el uso de técnicas de *machine learning*, una escala gravedad, comprensible y aplicable (*Spanish Influenza Score: SIS*), que permita categorizar o estratificar el riesgo de mortalidad en forma precoz, en los pacientes con gripe al ingreso en la UCI.

Objetivo primario

Desarrollar mediante técnicas de *machine learning*, una escala de estratificación de la gravedad (SIS) y evaluar la capacidad de predicción de mortalidad en UCI para pacientes con gripe grave en UCI

Objetivo secundario

Evaluar la capacidad de predicción de la mortalidad de un modelo no lineal como *Random Forest* (RF) en pacientes con gripe grave en UCI respecto de la escala SIS.

Material y método

Tipo de estudio

Subanálisis de base de datos GETGAG/SEMICYUC con pacientes ingresados en 184 UCI españolas en el periodo comprendido entre 2009-2018, con infección probada por gripe¹⁷.

Fuente de los datos

El conjunto de datos de entrenamiento (GE) y el de validación (GV) utilizados para el desarrollo del presente modelo, pertenecen a la base de datos creada en 2009 con motivo de la pandemia por la SEMICYUC para favorecer el conocimiento de esta nueva entidad y generar conocimientos de referencia para la mejor práctica clínica. El estudio fue aprobado por el Comité Ético de Investigación Clínica del Hospital Universitario Joan XXIII (IRB#11809) y ratificado por cada comité local de los centros participantes. La identidad de los pacientes permaneció anónima y no se solicitó consentimiento informado por la naturaleza observacional y epidemiológica del estudio, como ha sido publicado con anterioridad^{2,3,17-21}.

Se incluyeron todos los pacientes ingresados en forma consecutiva en las 184 UCI Españolas participantes, con signos clínicos respiratorios sugestivos de infección viral, con o sin fiebre y con confirmación microbiológica de gripe A o B mediante PCR-rt^{2,3,17-21}.

Los datos fueron obtenidos por los médicos asistenciales a través del examen físico, la revisión de la historia clínica, la radiología y los resultados de laboratorio. Los médicos asistenciales de cada centro fueron los responsables de solicitar todos los análisis y de realizar todos los procedimientos relacionados con el cuidado del paciente. Solamente se excluyeron del estudio pacientes menores de 15 años y aquellos pacientes con datos perdidos relacionados con los objetivos del estudio.

La base de datos contiene variables demográficas, nivel de gravedad, tiempo de inicio de los síntomas hasta el ingreso al hospital (Gap-H) o desde el ingreso al hospital hasta la UCI (Gap-UCI), o hasta el diagnóstico (Gap-Dg), vacunación antigripal, cuadrantes infiltrados en la radiografía de tórax, insuficiencia renal, ventilación no-invasiva (VNI) o invasiva (VM), fracaso a la VNI, shock al ingreso y comorbilidades, así como resultados laboratorio. La valoración de la severidad de la enfermedad se realizó a través del APACHE II score y el grado de disfunción orgánica a través del SOFA score.

Definiciones: las definiciones de las variables se pueden observar en la [tabla 1 del anexo B](#) y en publicaciones anteriores^{2,3,17-21}.

Tratamiento de valores perdidos

Se excluyeron los pacientes con datos perdidos en variables categóricas y se imputaron los valores perdidos de las variables numéricas mediante la función missForest/CRAN-R, una imputación no paramétrica del valor perdido utilizando árboles aleatorios^{22,23}.

Elección de los puntos de corte de las variables

Para el análisis, se debió convertir las variables numéricas continuas en categóricas. Los puntos de corte para las variables numéricas se obtuvieron de forma automática a través de la función LOESS *smoothing* (paquete stats/CRAN R). La regresión de LOESS²⁴ permite trazar curvas de una serie temporal mediante un método de regresión de mínimos cuadrados. Una vez obtenidas las curvas, se definen los puntos de corte a través de aquellas variaciones en la curva que se asocian con un aumento de al menos un 10% de la mortalidad.

Elección de las variables a incluir en el modelo

La elección de las variables se realizó de forma automática a través de la obtención del «information Value» para cada una de ellas mediante el paquete estadístico «InformationValue»-CRAN R. El *Information Value* (IV) es un método de búsqueda para seleccionar una variable predictora a través de regresión logística binaria^{25,26}. El IV total es la suma de los IV de su categoría y una medida de la capacidad predictiva de una variable, para discriminar «casos o eventos» y los «controles o no eventos». Para el IV se consideró un punto de corte $\geq 0,20$ para incluir las variables al modelo según lo sugeridos por Siddiqi²⁶.

Validación cruzada

En la [figura 1](#) se puede ver el flujo del análisis del estudio. La cohorte original de pacientes se dividió en dos grupos, GE (75% de los pacientes) para crear el modelo y GV (25% restante) para evaluar la precisión y el error de dicho modelo. La partición se realizó de forma aleatoria, pero manteniendo una misma proporción en la variable respuesta «y» (mortalidad).

Modelo de regresión y obtención de la puntuación del SIS

Una vez categorizadas todas las variables, se obtuvo un «valor» para cada nivel mediante un modelo binomial de regresión logística (RL) con la función «glm» de R. A partir de los coeficientes se obtuvieron los *ODDS RATIO* (OR), los cuales se redondearon para determinar los puntos asignados a cada variable del score SIS. Se aplicó la puntuación a cada uno de los pacientes y mediante la sumatoria, se obtuvo el puntaje final de la escala SIS. Se realizó dicho procedimiento para el GE y para GV y se evaluó la capacidad predictiva del modelo mediante su exactitud y discriminación mediante AUC ROC.

Conversión del puntaje en probabilidad de muerte y visualización de los resultados

Para obtener la probabilidad de muerte del puntaje SIS se estimaron mediante RL los coeficientes de la escala y la probabilidad del evento (mortalidad), utilizando los valores individuales de cada paciente. Posteriormente se construyó una figura de barras (*Barplot*) para representar los super-

Tabla 1 Características generales de los 3.959 pacientes incluidos en el presente análisis. Las variables son aquellas consideradas al ingreso a UCI hasta las primeras 24 horas de evolución (los resultados están expresado en número de pacientes (n) y porcentaje (%) o mediana y rango intercuartil (RIC) según corresponda. EPOC: enfermedad pulmonar obstructiva crónica; APACHE II: *Acute Physiology And Chronic Health Evaluation*; SOFA: *Sequential Organ Failure Assessment*; *Gap hospital*: tiempo desde inicio de los síntomas hasta ingreso en el Hospital; *Gap diagnóstico*: tiempo desde ingreso al Hospital y el diagnóstico; *Gap UCI*: tiempo desde el ingreso al hospital hasta la Unidad de Cuidados Intensivos-UCI; vacunados: pacientes que recibieron vacunación antigripal; IMC: índice de masa corporal)

Variables	Población total (n = 3.959)
Demográficas	
Edad	55 (43-67)
Sexo masculino	2.359 (59,6)
Tipo de diagnóstico de ingreso	
Neumonía viral primaria	2.520 (63,6)
Co-infección	805 (20,3)
EPOC reagudizado	280 (7,0)
Gravedad y nivel de asistencia	
APACHE II score	16 (12-21)
SOFA score	6 (4-8)
> 2 cuadrantes con infiltrados en la radiografía de tórax	1.731 (43,7)
Gap hospital	4 (2-6)
Gap diagnóstico	4 (2-7)
Gap UCI	1 (1-2)
Vacunados	466 (11,7)
Comorbilidades	
Asma	379 (9,6)
EPOC	938 (23,7)
Insuficiencia cardiaca crónica	531 (13,4)
Insuficiencia renal crónica	355 (8,9)
Enfermedad Hematológica	287 (7,2)
Embarazo	514 (12,9)
Obesidad (IMC >30)	1.239 (31,3)
Enfermedad neuromuscular	117 (2,9)
Enfermedad autoinmune	161 (4,0)
Inmunodeficiencia adquirida	445 (11,2)
Complicaciones	
Shock	
Ventilación mecánica invasiva	2.002 (50,5)
Ventilación mecánica no invasiva (VNI)	2.171 (54,8)
Fracaso de la VNI	1.455 (36,7)
Insuficiencia renal aguda	768 (19,4)
Mortalidad	1.129 (28,5)
	845 (21,3)

Modelo de regresión

Se dividió a la población en un grupo de entrenamiento (GE; n = 2.970) y en un grupo de validación (GV; n = 989). Las características de cada grupo pueden observarse en la [tabla 3 del anexo B](#). Los puntos de corte establecidos para APACHE II, SOFA y Gap-UCI y las variables categóricas ventilación VMI, *shock* e IRA se incluyeron en el modelo de regresión. En la [tabla 2](#) se pueden observar las variables asociadas de forma independiente con la mortalidad. Una vez aplicado el modelo en el GV, se observó una exactitud del 82% y un AUC ROC 82%.

Se transformaron los OR de cada variable en los puntos del score mediante redondeo al 0,5 más cercano y se

desarrolló un score con un puntaje máximo de 18 puntos ([tabla 3](#)). Se aplicó el puntaje a cada uno de los pacientes del GD y se obtuvo la predicción de mortalidad respecto de la puntuación para cada paciente ([fig. 2](#)). Después se aplicó el score al GV y se obtuvo una exactitud del 83% (95% IC: 0,79-0,84) con un área bajo la curva ROC del 82% ([fig. 3](#)) demostrando una buena discriminación del SIS. En la [figura 2 del anexo B](#) se puede observar que la calibración del modelo es buena con un índice de Sommers de 0,65 y en la [tabla 4](#) se observan los valores predictivos del SIS.

Finalmente, se determinaron cuatro niveles de riesgo de SIS estratificados según mortalidad: 1) Muy bajo riesgo: SIS 0-8.5 puntos con una mortalidad del 2,5%; 2) Riesgo moderado: SIS 9-11 puntos con una mortalidad del 16%; 3) Riesgo alto: SIS 11,5 -14 puntos con una mortalidad del 36.3% y 4)

Tabla 2 Variables asociadas de forma independiente con la mortalidad en UCI (análisis multivariado) (APACHE II: *Acute Physiology And Chronic Health Evaluation*; SOFA: *Sequential Organ Failure Assessment*; Gap UCI: tiempo desde el ingreso al hospital hasta la Unidad de Cuidados Intensivos)

Variable	OR	2,5% IC	97,5% IC	Valor de p =
<i>Intercept</i>	0,0157865	0,0096892	0,0251	< 1,1e-16***
Insuficiencia renal aguda	2,2759160	1,8238669	2,8398	3,247e-13***
Ventilación mecánica invasiva	3,7199974	2,8479344	4,8936	< 2,2e-16***
<i>Shock</i>	1,7920661	1,3835584	2,3270	1,078e-0,5***
APACHE II (11-17)	1,4155315	1,0114517	1,9993	0,0452860*
APACHE II (18-21)	1,9302566	1,3468351	2,7876	0,0003878***
APACHE II (22-27)	2,2490143	1,5392787	3,3074	3,203e-05***
APACHE II > 27 puntos	3,1892816	2,0516243	4,9825	2,924e-07***
SOFA (3-6)	1,1505601	0,8049519	1,6631	0,4478884
SOFA (7-8)	0,8934255	0,5981189	1,3436	0,5846590
SOFA (9-10)	1,3867558	0,9104261	2,1266	0,1303892
SOFA (11-12)	1,8822991	1,1519383	3,0919	0,0119529*
SOFA > 12 puntos	2,3234584	1,3389414	4,0551	0,0028335**
Gap-UCI (12-36)	1,4226239	1,0443490	1,9540	0,0272621*
Gap-UCI (37-60)	2,0356333	1,4050072	2,9611	0,0001834***
Gap-UCI (61-80)	3,2465693	2,0106550	5,2280	1,318e-06***
Gap-UCI >80 horas	4,3225489	3,0169562	6,2362	2-602e-15***

Significación *** 0.001; ** 0.01; *0.05

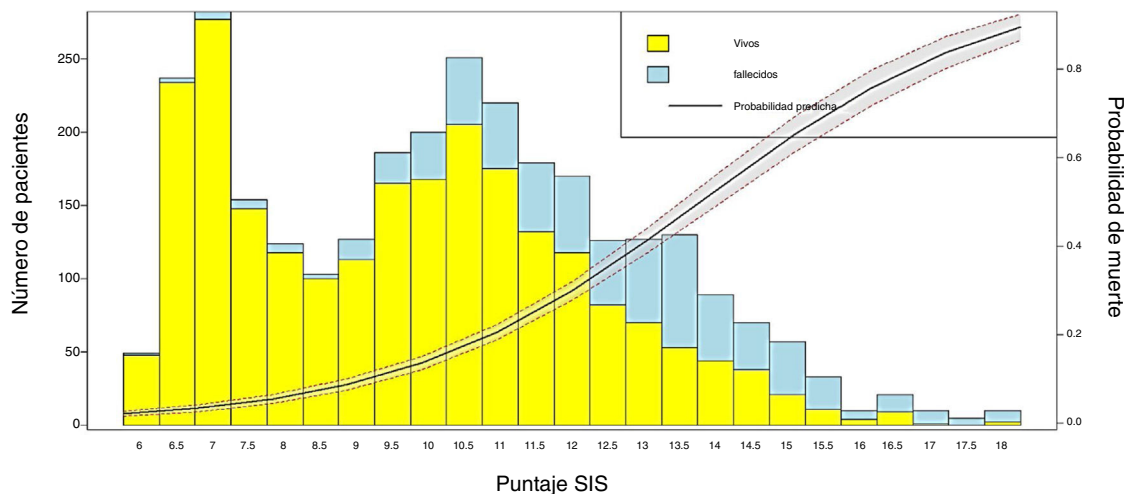


Figura 2 Gráfico de barras que muestra la aplicación de la puntuación obtenida SIS y la mortalidad observada, así como la curva de probabilidad de mortalidad para cada nivel. Se observa que la mortalidad se incrementa significativamente a medida que lo hace la puntuación obtenida ($p < 0,001$).

Muy alto riesgo: SIS > 14 puntos con una mortalidad del 60% (fig. 4).

82%. En la [tabla 4](#) se muestran el resto de los parámetros de predicción y su comparación respecto del SIS.

Modelo de predicción de mortalidad mediante Random Forest (modelo no lineal)

La aplicación de RF evidenció a la ventilación mecánica invasiva, SOFA score, insuficiencia renal aguda, días hasta la UCI, APACHE II, fracaso de la ventilación no invasiva e inmunodeficiencia como las variables con mayor impacto sobre la predicción (fig. 3 del anexo B). La valoración de la predicción evidenció una exactitud del 81% con un AUC ROC de

Discusión

La aplicación de escalas de gravedad a nivel individual y poblacional es de vital importancia pues permite clasificar y estratificar a los pacientes en categorías de riesgo, en función de uno de los resultados más importantes que pueden medirse en la UCI como es la mortalidad. Sobre este concepto, el objetivo principal de nuestro estudio fue desarrollar un modelo de predicción «precoz» de mortalidad, mediante aplicación de métodos de *machine learning* y

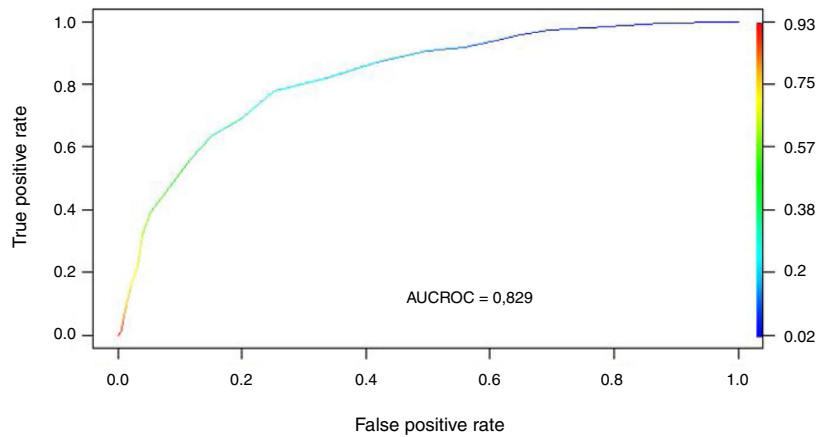


Figura 3 Área bajo la curva ROC (AUCROC) para la puntuación SIS obtenida a través en el grupo de validación.

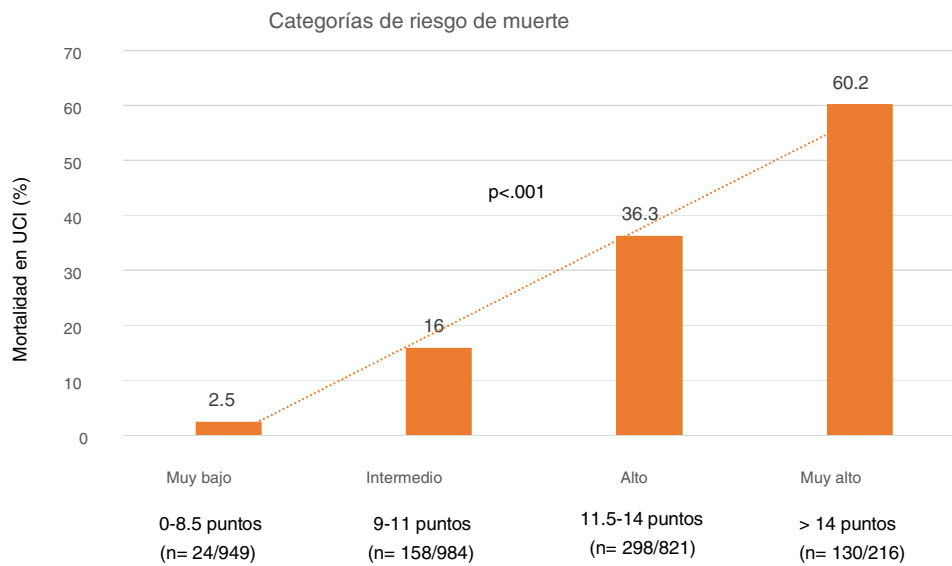


Figura 4 Categorías de riesgo de mortalidad estratificadas según el Spanish Influenza Score (SIS) donde se puede observar como la mortalidad aumenta de forma significativa a medida que lo hace el riesgo.

comparar su rendimiento respecto de un modelo no lineal de Random Forest.

El principal hallazgo del estudio fue que la escala SIS desarrollada, demostró una adecuada exactitud en la validación cruzada (83%) y una muy buena discriminación (AUCROC 82%), parámetros de predicción similares al modelo Random Forest. Estos datos sugieren que la escala SIS es un modelo válido, que permite estratificar adecuadamente el riesgo de muerte en pacientes con gripe al ingreso a UCI.

Hasta el momento los estudios realizados solo han determinado variables asociadas a la mortalidad a través de análisis multivariado clásico^{5,6,19,30} o bien, desarrollando escalas con un limitado número de pacientes^{7,31-33} o considerando solamente poblaciones especiales^{9,10}. Oh et al.⁷ en 709 pacientes, desarrollaron una escala con 4 variables, adjudicando un punto para cada una de ellas (estado mental alterado, hipoxia, infiltrados bilaterales y edad > 65 años). Aunque es un estudio multicéntrico y la discriminación de la escala fue muy buena (AUC ROC 0,83), solo 75 (10,5%) pacientes fueron casos graves. Además, los autores no

realizaron validación cruzada. Adeniji et al.⁸ aplicaron el STSS (*Simple Triage Scoring System*)³¹ y el SOFA en urgencias, para predecir la necesidad de ventilación mecánica (VM) y el ingreso en UCI de pacientes con gripe. La discriminación fue superior para el STSS (AUC ROC 0,88) respecto del SOFA (AUC ROC 0,77) para el ingreso a UCI y también para necesidad de VM (AUC ROC 0,91 vs. AUC ROC 0,87 para STSS y SOFA respectivamente). Sin embargo, el número de pacientes incluidos es muy bajo (n = 62), por lo cual la potencia del estudio es pobre y los resultados difíciles de interpretar. Chung et al.⁹ desarrollaron una escala de gravedad en 409 pacientes ancianos (GID: Geriatric Influenza Death). El análisis multivariado evidenció solo cinco variables (coma, elevación de la PCR, cáncer, enfermedad coronaria y presencia de bandas en la fórmula leucocitaria) asociadas de forma independiente con la mortalidad. Aunque el GID, mostró una muy buena discriminación (AUC ROC 86%), a diferencia de nuestra escala, considera variables de toda la evolución, se limita a la población de mayores y no se ha realizado validación cruzada.

Tabla 3 Puntaje del *Spanish Influenza Score* (SIS) derivado de los OR de la regresión logística (IRA: Insuficiencia renal aguda; VMI: ventilación mecánica invasiva; APACHE II: *Acute Physiology And Chronic Health Evaluation*; SOFA: *Sequential Organ Failure Assessment*; Gap UCI: tiempo desde el ingreso al hospital hasta la Unidad de Cuidados Intensivos)

Variable	Puntos
Presencia IRA	2.5
Necesidad VMI	3.5
Presencia shock	2.0
APACHE (puntos)	
11-17	1.5
18-21	2.0
22-27	2.0
>27	3.0
SOFA (puntos)	
3-6	1.0
7-8	1.0
9-10	1.5
11-12	2.0
>12	2.5
GAPUCI (hs)	
12-36	1.5
37-60	2.0
61-80	3.0
>80	4.5
Puntuación máxima	18.0

Tabla 4 Valores predictivos del *Spanish Influenza Score* (SIS) y del modelo mediante *Random Forest* (RF) para los 3.959 pacientes incluidos

Variables	Modelo SIS	Modelo RF
Exactitud (<i>accuracy</i>)	83%	81%
Sensibilidad	93.7%	95.7
Especificidad	38.4%	30.3
Valor predictivo positivo	84.0	83.4
Valor predictivo negativo	62.0	64.0
AUC ROC	82%	82%

Los estudios basados en métodos estadísticos habituales, como la regresión logística (modelo lineal), son ampliamente aceptados por los médicos para determinar o investigar factores relacionados con la mortalidad o el desarrollo de algún evento adverso. Sin embargo, estos indicadores no funcionan adecuadamente para la predicción individual¹⁵, ni permiten predecir el rumbo clínico de un paciente. Nuevas formas de predicción, utilizando algoritmos desarrollados mediante técnicas de aprendizaje automático (*machine learning*), como las redes neuronales o los árboles de decisión, han sido implementados para la obtención de modelos de predicción en diferentes escenarios en cuidados intensivos³⁴⁻³⁷. Pero estos modelos, aunque demuestran muy buen rendimiento en la predicción, suelen ser incomprensibles para los clínicos y poco aplicables, no solo por su complejidad, sino también, por la falta de inclusión en el modelo de variables de gran interés clínico, como

el tratamiento antimicrobiano, en un modelo complejo que compara modelos clínicos versus automáticos en el tratamiento de la sepsis³⁸, lo cual invalida la aplicación clínica del modelo. De forma reciente, Hu et al.³⁹ publicaron un estudio sobre la aplicación de dos técnicas de *machine learning* (*gradient boosting XGBoost* y RF) comparadas con un modelo de RL para predecir la mortalidad a los 30 días, en una cohorte de 336 pacientes con gripe. Los autores concluyen que los modelos XGBoost (AUC ROC = 0,84) y RF (AUC ROC = 0,80) presentaron una mejor discriminación respecto de la RL (AUC ROC = 0,70). Estos resultados no se corresponden con lo observado en nuestro estudio, donde también utilizamos un modelo de RF. La discrepancia podría estar relacionada con el bajo número de pacientes respecto de un elevado número de variables consideradas en el estudio de Hu et al.³⁹, lo cual afecta de forma desfavorable a los modelos de regresión, pero no a los desarrollados mediante árboles de decisión. Además, utiliza variables de los primeros 7 días, con lo cual no puede considerarse como un predictor precoz. Finalmente, la discriminación del mejor modelo (XGBoost), el cual es poco interpretable para el clínico, solo es ligeramente superior a la observada en la escala SIS.

Las escalas utilizadas de forma habitual en UCI para medir la gravedad general (APACHE II) o el grado de disfunción orgánica (SOFA) presentan limitaciones al momento de categorizar los pacientes con gripe grave^{7,8,39-41}, por lo cual nuestra escala SIS, podría ser una alternativa sencilla a implementar en este grupo de pacientes, ya que ha demostrado un rendimiento similar a un modelo de predicción desarrollado con *Random Forest* (RF). Aunque el RF es uno de los mejores métodos de *machine learning* para responder a los problemas complejos, especialmente aquellos relacionados con asociaciones no lineales²⁹, la principal desventaja del método es que es poco entendible para el clínico, ya que no permite conocer cómo se realizan las asociaciones (cajas negras) para generar el modelo final. En el mismo sentido que nuestros resultados, pero en la población general de UCI, Kim et al.¹⁵ estudiaron la capacidad de predicción de mortalidad, de tres diferentes modelos desarrollados con *machine learning* (redes neuronales, *support vector machine* y árboles de decisión), respecto de un modelo tradicional de regresión logística desarrollado con las variables del APACHE III score⁴². El estudio incluyó a más de 38.000 ingresos y solo consideró los datos capturados en las primeras 24 horas de ingreso a UCI. El estudio evidenció una capacidad de predicción similar entre los cuatro modelos, resaltando a la regresión logística como un método válido para determinar la predicción de mortalidad respecto de modelos más complejos.

Nuestro estudio combina técnicas de aprendizaje automático con regresión logística, lo cual le brinda robustez y objetividad. Además, siendo un estudio multicéntrico y con un elevado número de pacientes, permite la generalización de los resultados, ya que las 184 UCI participantes representan aproximadamente el 50% de todas las UCI de España. Sin embargo, nuestro estudio presenta algunas limitaciones que deben ser remarcadas para su adecuada interpretación. Primero, nuestra escala SIS utiliza solamente datos al ingreso de UCI, por lo cual, datos relacionados con los cambios producidos durante la evolución no son considerados. Aunque esto puede afectar la capacidad de predicción, nuestro

objetivo primario fue desarrollar una escala de riesgo «precoz» al momento de ingreso a UCI y no durante su evolución, la cual ha demostrado un adecuado poder de discriminación. Segundo, el desarrollo del modelo se ha realizado considerando solamente pacientes ingresados en UCI españolas, por lo cual, el modelo podría no presentar un adecuado funcionamiento en otros países, o en otras poblaciones fuera del ámbito de la UCI. Tercero, aunque se ha realizado una validación cruzada, el funcionamiento de SIS no ha sido probado en forma prospectiva, por lo cual, nuestro proyecto contempla una validación prospectiva nacional e internacional del SIS para evaluar el impacto clínico real y la aceptación de la escala por parte de los intensivistas.

En conclusión, la escala SIS desarrollada en base a los datos de más de 3.900 pacientes críticos, demuestra una predicción similar a la observada para un modelo de Random Forest. Considerando que la aplicabilidad de la escala SIS es sencilla y permite obtener una estratificación del riesgo de muerte de forma precoz, su implementación podría impactar favorablemente en la evolución de los pacientes ingresados en UCI por infección grave por gripe. Sin embargo, estas consideraciones deberán ser confirmadas con la validación prospectiva del SIS.

Autoría/colaboraciones

Concepción y el diseño del estudio: AR (Alejandro Rodríguez), ED (Emili Díaz) ST(Sandra Trefler), JMC (Judth Marín-Corral), LC(Laura Claverias). IML(Ignacio Martín Loeches),MB(María Bodi), JSV(Jordi Sole-Violan), JG(Jose garnacho-Montero), MRB (Manuel Ruiz Botella), JG (Josep Gomez), JA(Jordi Albiol), EM (Eduard mallol)

Adquisición de los datos y análisis: AR (Alejandro Rodríguez), ST(Sandra Trefler), LC(Laura Claverias), GM(Gerard Moreno), MS (Manuel Samper), MB(María Bodi), JMC (Judth Marín-Corral), MRB (Manuel Ruiz Botella), JG (Josep Gomez), JA(Jordi Albiol), EM (Eduard mallol) AB(Ariel Barrios)

Interpretación de los datos: AR (Alejandro Rodríguez), MS (Manuel Samper), GM(Gerard Moreno), MB(María Bodi), ED (Emili Díaz) ST(Sandra Trefler), JMC (Judth Marín-Corral), LC(Laura Claverias), JCY (Juan carlos yebenes), AT (Antoni Torres) PR(Paula Ramirez) JGM(Jose Garnacho Montero), RF(Ricard ferrer), IML(Ignacio Martín Loeches, LFR(Luis Felipe Reyes) JG(JUAN Guardiola), MIR(Marcos I restrepo), JSV(Jordi Sole-Violan)

Importante aportación intelectual al contenido: AT (Antoni Torres), JGM (Jose Garnacho Montero), PR (Paula Ramirez), JCY (Juan Carlos Yebenes), RF (Ricard ferrer), LFR(Luis Felipe Reyes), JG (JUAN Guardiola),MIR(Marcos I Restrepo)

Borrador del manuscrito: AR (Alejandro Rodríguez), MS (Manuel Samper), GM(Gerard Moreno)

Revisión crítica del contenido: MB(María Bodi), ED (Emili Díaz). JMC (Judth Marín-Corral), LC(Laura Claverias), JCY (Juan carlos yebenes), JSV(Jordi Sole-Violan), AT (Antoni Torres), PR (Paula Ramirez), JGM (Jose Garnacho Montero), RF (Ricard ferrer), IML(Ignacio Martín Loeches), LFR(Luis Felipe Reyes), AB(Ariel Barrios), JG(JUAN Guardiola), MIR (Marcos I restrepo)

Todos los autores aprobaron el manuscrito final enviado para su evaluación y posible publicación.

Los hallazgos y las conclusiones del presente manuscrito son responsabilidad de los autores y no representan necesariamente la posición oficial de la SEMICYUC.

Papel de los auspiciantes

La SEMICYUC no ha tenido ninguna implicación en el diseño del estudio, en el análisis y la interpretación de los datos o en la redacción del presente manuscrito. JG ha tenido un tiempo parcial protegido destinado al análisis del estudio, a través de una beca de investigación de la Fundación Privada Barri. AR, el autor para la correspondencia, ha tenido acceso a todos los datos del estudio y es el responsable final de enviar el manuscrito para su publicación.

Conflicto de intereses

AR ha recibido una beca de investigación de Gilead Science para el estudio de antibióticos nebulizados. Además, ha recibido honorarios por conferencias docentes para Biomeirieux, Astellas, Pfizer, Thermo Fisher, MSD, Gilead, Shionogi y BRHAMS. Sin embargo, no tiene conflictos de interés con el presente manuscrito

Agradecimientos

Este estudio está auspiciado por la SEMICYUC (Sociedad Española de Medicina Intensiva, Crítica y Unidades Coronarias). Los autores agradecen a todos los investigadores del GETGAG (Grupo Español de Trabajo Gripe A Grave) por su continua participación en el proyecto desde el año 2009, sin la cual este proyecto no podría haberse realizado.

ANEXO. AUTORES

Coordinadores clínicos: M. Samper, G. Moreno, M. Bodi, E. Díaz, J. Marín-Corral, L. Claverias, S. Trefler, J.C. Yebenes, J. Solé Violán, A. Torres, P. Ramírez, J. Garnacho-Montero, R. Ferrer, A. Rodríguez

Coordinadores científicos de datos: M. Ruiz-Botella, J. Gómez, J. Albiol, E. Mayol

Consultores externos: I. Martín-Loeches, L.F. Reyes, A. Barrios, J. Guardiola, M.I. Restrepo

Anexo A. Material adicional

Se puede consultar material adicional a este artículo en su versión electrónica disponible en [doi:10.1016/j.medin.2020.05.017](https://doi.org/10.1016/j.medin.2020.05.017).

Bibliografía

- Rodríguez A, Martín-Loeches I, Bonastre J, Olaechea P, Alvarez-Lerma F, Zaragoza R, et al. Primera epidemia de gripe estacional después de la pandemia por gripe A en 2009: Descripción de los primeros 300 ingresos en UCI españolas. *Med Intensiva*. 2011;35:208–16, <http://dx.doi.org/10.1016/j.medin.2011.03.001>.
- Martin-Loeches I, Diaz E, Vidaur L, Torres A, Laborda C, Granada R, et al. Pandemic and post-pandemic Influenza A (H1N1)

- infection in critically ill patients. *Crit Care*. 2011;15:R286, <http://dx.doi.org/10.1186/cc10573>.
3. Martin-Loeches I, Rodríguez A, Bonastre J, Zaragoza R, Sierra R, Marques A, et al. Severe pandemic (H1N1)v influenza A infection: Report on the first deaths in Spain. *Respirology*. 2011;16:78–85, <http://dx.doi.org/10.1111/j.1440-1843.2010.01874.x>.
 4. Rodríguez A, Ferri C, Martin-Loeches I, Díaz E, Masclans J, Gordo F, et al. Risk Factors for Noninvasive Ventilation Failure in Critically Ill Subjects With Confirmed Influenza Infection. *Respir Care*. 2017;62:1307–15, <http://dx.doi.org/10.4187/respcare.05481>.
 5. Shi SJ, Li H, Liu M, Liu YM, Zhou F, Liu B, et al. Mortality prediction to hospitalized patients with influenza pneumonia: PO2/FiO2 combined lymphocyte count is the answer. *Clin Respir J*. 2017;11:352–60, <http://dx.doi.org/10.1111/crj.12346>.
 6. Carpenter CR, Keim SM, Upadhye S, Nguyen HB. Risk Stratification of the Potentially Septic Patient in the Emergency Department: The Mortality in the Emergency Department Sepsis (MEDS) Score. *J Emerg Med*. 2009;37:319–27, <http://dx.doi.org/10.1016/j.jemermed.2009.03.016>.
 7. Oh WS, Lee SJ, Lee CS, Hur JA, Hur AC, Park YS, et al. A prediction rule to identify severe cases among adult patients hospitalized with pandemic influenza a (H1N1) 2009. *J Korean Med Sci*. 2011;26:499–506, <http://dx.doi.org/10.3346/jkms.2011.26.4.499>.
 8. Adeniji KA, Cusack R. The Simple Triage Scoring System (STSS) successfully predicts mortality and critical care resource utilization in H1N1 pandemic flu: A retrospective analysis. *Crit Care*. 2011;15:R39, <http://dx.doi.org/10.1186/cc10001>.
 9. Chung JY, Hsu CC, Chen JH, Chen WL, Lin HJ, Guo HR, et al. Geriatric influenza death (GID) score: A new tool for predicting mortality in older people with influenza in the emergency department. *Sci Rep*. 2018;8:1–8, <http://dx.doi.org/10.1038/s41598-018-27694-6>.
 10. Pappalardo F, Pieri M, Greco T, Patroniti N, Pesenti A, Arcadipane A, et al. Predicting mortality risk in patients undergoing venovenous ECMO for ARDS due to influenza A (H1N1) pneumonia: The ECMOnet score. *Intensive Care Med*. 2013;39:275–81, <http://dx.doi.org/10.1007/s00134-012-2747-1>.
 11. Deliberato RO, Escudero GG, Bulgarelli L, Neto AS, Ko SQ, Campos NS, et al. SEVERITAS: An externally validated mortality prediction for critically ill patients in low and middle-income countries. *Int J Med Inform*. 2019;131.
 12. Kim SY, Kim S, Cho J, Kim YS, Sol IS, Sung Y, et al. A deep learning model for real-time mortality prediction in critically ill children. *Crit Care*. 2019;23:1–10, <http://dx.doi.org/10.1186/s13054-019-2561-z>.
 13. Meyer A, Zverinski D, Pfahring B, Kempfert J, Kuehne PT, Sündermann SH, et al. Machine learning for real-time prediction of complications in critical care: a retrospective study. *Lancet Respir Med*. 2018;6:905–14, [http://dx.doi.org/10.1016/S2213-2600\(18\)30300-X](http://dx.doi.org/10.1016/S2213-2600(18)30300-X).
 14. Anand RS, Stey P, Jain S, Biron DR, Bhatt H, Monteiro K, et al. Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci*. 2018;2017:310–9. Disponible en <http://www.ncbi.nlm.nih.gov/pubmed/29888089> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5961793>.
 15. Kim S, Kim W, Woong Park R. A comparison of intensive care unit mortality prediction models through the use of data mining techniques. *Healthc Inform Res*. 2011;17:232–43, <http://dx.doi.org/10.4258/hir.2011.17.4.232>.
 16. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;9:15–8, <http://dx.doi.org/10.1038/s41591-019-0548-6>.
 17. Moreno G, Rodríguez A, Reyes LF, Gomez J, Sole-Violan J, Díaz E, et al. Corticosteroid treatment in critically ill patients with severe influenza pneumonia: a propensity score matching study. *Intensive Care Med*. 2018;44:1470–82, <http://dx.doi.org/10.1007/s00134-018-5332-4>.
 18. Garnacho-Montero J, Gutiérrez-Pizarra A, Márquez JA, Zaragoza R, Granada R, Ruiz-Santana S, et al. Epidemiology, clinical features, and prognosis of elderly adults with severe forms of influenza A (H1N1). *J Am Geriatr Soc*. 2013;61:350–6, <http://dx.doi.org/10.1111/jgs.12152>.
 19. Álvarez-Lerma F, Marín-Corral J, Vila C, Masclans JR, Molina FJ, Loeches IM, et al. Delay in diagnosis of influenza A (H1N1)pdm09 virus infection in critically ill patients and impact on clinical outcome. *Crit Care*. 2016;20:337, <http://dx.doi.org/10.1186/s13054-016-1512-1>.
 20. Alvarez-Lerma F, Marín-Corral J, Vilá C, Masclans JR, Loeches IM, Barbadillo S, et al. Characteristics of patients with hospital-acquired influenza A (H1N1)pdm09 virus admitted to the intensive care unit. *J Hosp Infect*. 2017;95:200–6, <http://dx.doi.org/10.1016/j.jhin.2016.12.017>.
 21. Garnacho-Montero J, León-Moya C, Gutiérrez-Pizarra A, Arenzana-Seisdedos A, Vidaur L, Guerrero JE, et al. Clinical characteristics, evolution, and treatment-related risk factors for mortality among immunosuppressed patients with influenza A (H1N1) virus admitted to the intensive care unit. *J Crit Care*. 2018;48:172–7, <http://dx.doi.org/10.1016/j.jcrr.2018.08.017>.
 22. Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28:112–8, <http://dx.doi.org/10.1093/bioinformatics/btr597>.
 23. Van HC, Sauerbrei W. Cross-Validation, Shrinkage and Variable Selection in Linear Regression Revisited. *Open J Stat*. 2013;3:79–102, <http://dx.doi.org/10.4236/ojs.2013.32011>.
 24. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc*. 1979;74:829–36, <http://dx.doi.org/10.1080/01621459.1979.10481038>.
 25. Zhang Z, Zhang H, Khanal MK. Development of scoring system for risk stratification in clinical medicine: A step-by-step tutorial. *Ann Transl Med*. 2017;5:1–9, <http://dx.doi.org/10.21037/atm.2017.08.22>.
 26. Siddiqi N. Scorecard Development Process. En: Siddiqi N, editor. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. New Jersey: John Wiley & Sons; 2015. p. 73–127.
 27. Sommers R. A New Asymmetric Measure of Association for Ordinal Variables. *Am Sociol Rev*. 1962;27:799–811.
 28. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. *Front Aging Neurosci*. 2017;9:1–12, <http://dx.doi.org/10.3389/fnagi.2017.00329>.
 29. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32, <http://dx.doi.org/10.1201/9780429469275-8>.
 30. Teng F, Wan TT, Guo SB, Liu J, Cai JF, Qi X, et al. Outcome prediction using the Mortality in Emergency Department Sepsis score combined with procalcitonin for influenza patients. *Med Clin*. 2019;153:411–7, <http://dx.doi.org/10.1016/j.medcli.2019.03.022>.
 31. Talmor D, Jones AE, Rubinson L, Howell MD, Shapiro NI. Simple triage scoring system predicting death and the need for critical care resources for use during epidemics. *Crit Care Med*. 2007;35:1251–6, <http://dx.doi.org/10.1097/01.CCM.0000262385.95721.CC>.
 32. Hak E, Wei F, Nordin J, Mulla J, Poblete S, Nichol KL. Development and Validation of a Clinical Prediction Rule for Hospitalization Due to Pneumonia or Influenza or Death during Influenza Epidemics among Community-Dwelling Elderly Persons. *J Infect Dis*. 2004;189:450–8, <http://dx.doi.org/10.1086/381165>.

33. Moa A, Muscatello D, Chughtai A, Chen X, Raina MacIntyre C. Flucast: A real-time tool to predict severity of an influenza season. *J Med Internet Res*. 2019;21:e11780, <http://dx.doi.org/10.2196/11780>.
34. Morik K, Imboff M, Brockhausen P, Joachims T, Gather U. Knowledge discovery and knowledge validation in intensive care. *Artif Intell Med*. 2000;19:225–49, [http://dx.doi.org/10.1016/S0933-3657\(00\)00047-6](http://dx.doi.org/10.1016/S0933-3657(00)00047-6).
35. Moser SA, Jones WT, Brossette SE. Application of data mining to intensive care unit microbiologic data. *Emerg Infect Dis*. 1999;5:454–7, <http://dx.doi.org/10.3201/eid0503.990320>.
36. Ganzert S, Guttman J, Kersting K, Kuhlen R, Putensen C, Sydow M, et al. Analysis of respiratory pressure-volume curves in intensive care medicine using inductive machine learning. *Artif Intell Med*. 2002;26(1-2):69–86, [http://dx.doi.org/10.1016/s0933-3657\(02\)00053-2](http://dx.doi.org/10.1016/s0933-3657(02)00053-2).
37. Kong L, Milbrandt EB, Weissfeld LA. Advances in statistical methodology and their application in critical care. *Curr Opin Crit Care*. 2004;10:391–4, <http://dx.doi.org/10.1097/01.ccx.0000140940.96505.71>.
38. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*. 2018;24:1716–20, <http://dx.doi.org/10.1038/s41591-018-0213-5>.
39. Hu CA, Chen CM, Fang YC, Liang SJ, Wang HC, Fang WF, et al. Using a machine learning approach to predict mortality in critically ill influenza patients: a cross-sectional retrospective multicentre study in Taiwan. *BMJ Open*. 2020;10:e033898, <http://dx.doi.org/10.1136/bmjopen-2019-033898>.
40. Khan Z, Hulme J, Sherwood N. An assessment of the validity of SOFA score based triage in H1N1 critically ill patients during an influenza pandemic. *Anaesthesia*. 2009;64:1283–8, <http://dx.doi.org/10.1111/j.1365-2044.2009.06135.x>.
41. Guest T, Tantam G, Donlin N, Tantam K, McMillan H, Tillyard A. An observational cohort study of triage for critical care provision during pandemic influenza: «Clipboard physicians» or «evidenced based medicine»? *Anaesthesia*. 2009;64:1199–206, <http://dx.doi.org/10.1111/j.1365-2044.2009.06084.x>.
42. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system: Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest*. 1991;100:1619–36, <http://dx.doi.org/10.1378/chest.100.6.1619>.