



Programa de Doctorado en Salud Pública, Ciencias Médicas y
Quirúrgicas

**REVISIÓN SISTEMÁTICA DE LA CALIDAD DE LOS MODELOS
PREDICTIVOS DE MORTALIDAD EN SEPSIS**

Concepción Beneyto Ripoll

Directora de la tesis

Dra. Dña. Maria Ángeles Carbonell Torregrosa

Codirector de la tesis

Dr. D. Antonio Palazón Bru

Universidad Miguel Hernández de Elche

2023





La presente Tesis Doctoral, titulada “REVISIÓN SISTEMÁTICA DE LA CALIDAD DE LOS MODELOS PREDICTIVOS DE MORTALIDAD EN SEPSIS”, se presenta bajo la modalidad de **tesis convencional con el siguiente indicio de calidad:**

- Beneyto-Ripoll C, Palazón-Bru A, Llópez-Espinós P, Martínez-Díaz AM, Gil-Guillén VF, de los Ángeles Carbonell-Torregrosa M. A critical appraisal of the prognostic predictive models for patients with sepsis: Which model can be applied in clinical practice?. Int J Clin Pract. 2021;00:e14044. <https://doi.org/10.1111/ijcp.14044>
- International Journal of Clinical Practice es una revista situada en el año 2021 en el segundo cuartil, en la categoría científica Medicina General e Interna, con un factor de impacto de 3,149, en el «Journal Citation Reports (JCR) Science Edition.





La Dra. Dña. *“Maria Ángeles Carbonell Torregrosa”*, directora, y el Dr. D. *“Antonio Palazón Bru”*, codirector de la tesis doctoral titulada **“Revisión sistemática de la calidad de los modelos predictivos de mortalidad es sepsis”**

INFORMAN:

Que D./Dña. *“Concepción Beneyto Ripoll”* ha realizado bajo nuestra supervisión el trabajo titulado **“Revisión sistemática de la calidad de los modelos predictivos de mortalidad es sepsis”** conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

En Alicante, ade de 2023

Directora de la tesis

Dra. Dña. *“Maria Ángeles Carbonell Torregrosa”*

Codirector de la tesis

Dr. D. *“Antonio Palazón Bru”*





La Dra. Dña. *"María del Mar Masiá Canuto"*, Coordinadora del Programa de Doctorado en **Salud Pública, Ciencias Médicas y Quirúrgicas**.

INFORMA:

Que D./Dña. *"Concepción Beneyto Ripoll"* ha realizado bajo la supervisión de nuestro Programa de Doctorado el trabajo titulado **"Revisión sistemática de la calidad de los modelos predictivos de mortalidad es sepsis"** conforme a los términos y condiciones definidos en su Plan de Investigación y de acuerdo al Código de Buenas Prácticas de la Universidad Miguel Hernández de Elche, cumpliendo los objetivos previstos de forma satisfactoria para su defensa pública como tesis doctoral.

Lo que firmo para los efectos oportunos, en a de de 2023

Profa. Dra. Dña. *"María del Mar Masiá Canuto"*

Coordinadora del Programa de Doctorado en Salud Pública, Ciencias Médicas y Quirúrgicas.



ÍNDICE

1.	LISTADO DE ABREVIATURAS.....	11
2.	LISTADO DE FIGURAS Y TABLAS.....	15
3.	RESUMEN EN CASTELLANO	19
4.	RESUMEN EN INGLÉS.....	23
5.	INTRODUCCIÓN.....	27
5.1.	Sepsis.....	29
5.1.1.	Antecedentes y epidemiología.....	29
5.1.2.	Definición de sepsis y otros conceptos.....	30
5.1.3.	Recomendaciones en el tratamiento de la sepsis.....	37
5.1.3.1	Detección precoz de sepsis	38
5.1.3.2.	Tratamiento de la sepsis	43
5.1.4.	Factores de riesgo de infección, sepsis, fallo orgánico y mortalidad.	45
5.2.	Modelos predictivos.....	49
5.2.1.	Definición.....	50
5.2.2.	Utilidades y limitaciones.....	50
5.2.3.	Calidad de los modelos pronósticos.....	52
5.2.4.	Directrices internacionales.....	54
6.	JUSTIFICACIÓN.....	63
7.	OBJETIVOS	67
8.	MATERIALES Y MÉTODOS.....	71
8.1.	Protocolo y registro.....	73
8.2.	Criterios de inclusión y exclusión.....	73
8.3.	Fuente de datos	74
8.4.	Búsqueda y selección de estudios	74
8.5.	Extracción de datos	75
8.6.	Análisis de la calidad.....	76
8.7.	Análisis estadístico	76
9.	RESULTADOS.....	79
10.	DISCUSIÓN.....	93
10.1.	Fortalezas y limitaciones.....	95
10.2.	Comparación con la literatura existente.....	95
10.3.	Implicaciones en la investigación y en la práctica clínica.....	96

11.	CONCLUSIÓN	101
12.	REFERENCIAS	105
13.	ANEXO.....	117
14.	AGRADECIMIENTOS.....	151



1.LISTADO DE ABREVIATURAS





ACCP: American College of Chest Physicians

APACHE: Acute Physiology and Chronic Health

ATS: American Thoracic Society

BUN: nitrógeno ureico

CHARMS: CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies

CONSORT: Consolidated Standards of Reporting Trials

CSS: Campaña Sobrevivir a la Sepsis

EPV: razón eventos por variable

ESICM: European Society of Intensive Care Medicine

FC: frecuencia cardíaca

FiO₂: fracción de oxígeno inspirado

FR: frecuencia respiratoria

GRIPS: Genetic Risk Prediction Studies

LODS: Logistic Organ Dysfunction System

MEWS: Modified Early Warning Score

MODS: Multiple Organ Dysfunction Score

MPM: Mortality Probability Model

NEWS: National Early Warning Score

PaO₂: presión arterial de oxígeno

PaO₂/FiO₂: presión arterial de oxígeno/fracción de oxígeno inspirado

pCO₂: presión parcial de dióxido de carbono

PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

PROBAST: Prediction model Risk Of Bias ASsessment Tool

qSOFA: quick Sequential Organ Failure Assessment

QUADAS: Quality Assessment of Diagnostic Accuracy Studies

QUIPS: Quality In Prognosis Studies Tool

REMARK: REporting recommendations for tumour MARKer prognostic studies

ROC: receiver operating characteristic

SAPS: Simplified Acute Physiology Score

SCCM: Society of Critical Care Medicine

SIS: Surgical Infection Society

SOFA: Sequential Organ Failure Assessment

SRIS: Síndrome de respuesta inflamatoria sistémica

STARD: Statement for Reporting Studies of Diagnostic Accuracy

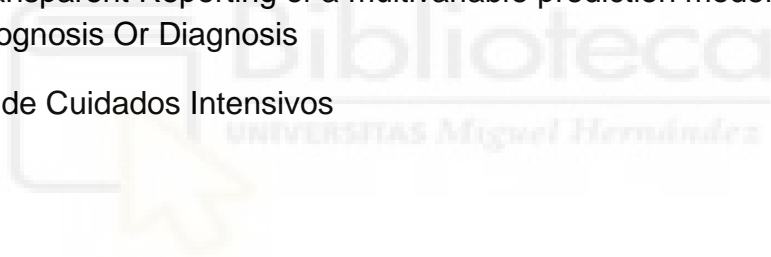
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

TAM: tensión arterial media

TAS: tensión arterial sistólica

TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

UCI: Unidad de Cuidados Intensivos



2.LISTADO DE FIGURAS Y TABLAS





Tabla 1.- Criterios diagnósticos de sepsis 2001.....	32
Tabla 2.- Escala de medición de disfunción orgánica: <i>Sequential Organ Failure Assessment</i> (SOFA).....	36
Tabla 3.- Escala quickSOFA.....	39
Tabla 4.- National Early Warning Score (NEWS).....	41
Tabla 5.- Modified Early Warning Score (MEWS).....	41
Tabla 6.- Factores de riesgo relacionados con las infecciones.....	45
Tabla 7.- Factores relacionados con disfunción orgánica.....	47
Tabla 8.- Escalas de gravedad en pacientes críticos.....	48
Tabla 9.- Discriminación y calibración de los modelos predictivos incluidos en nuestra revisión sistemática.....	84
Tabla 10.- Resumen resultados del análisis CHARMS.....	87
Tabla 11.- Riesgo de sesgo y preocupación respecto a la aplicabilidad de los estudios incluidos que evalúan modelos pronósticos multivariados que predicen mortalidad en pacientes con sepsis. (PROBAST).....	90
Tabla 12.- Bajo riesgo de sesgo y baja preocupación respecto a la aplicabilidad (%) de otras revisiones sistemáticas que evalúan modelos predictivos multivariados para otras enfermedades. (PROBAST).....	91
Tabla 13.- Resumen apartado Discusión.....	99
Figura 1.- Esquema de la revisión sistemática.....	81



3.RESUMEN EN CASTELLANO





La sepsis es un trastorno orgánico secundario a una infección que tiene una elevada incidencia, constituyendo una de las principales causas de morbimortalidad en todo el mundo. Su identificación y manejo precoz son claves para mejorar la supervivencia, especialmente en los pacientes más graves.

La estratificación de la gravedad de la sepsis mediante modelos de riesgo es primordial para adecuar el tratamiento con el fin de obtener unos mejores resultados. Existen multitud de modelos pronósticos, pero hay pocos implementados en la práctica clínica. Su uso es recomendable, pero puesto que van a tener un impacto clínico, es necesario que estos modelos estén desarrollados según las recomendaciones internacionales y que hayan sido validados en otras poblaciones. Además, es importante evaluar su aplicabilidad o potencial uso en la práctica clínica habitual.

Nos propusimos realizar una revisión sistemática de modelos predictivos de mortalidad en pacientes con sepsis, acorde a las recomendaciones internacionales para este tipo de revisiones, como son las recomendaciones CHARMS (Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies), y la guía PROBAST (Prediction model Risk Of Bias ASsessment Tool), con el objetivo de obtener un documento crítico que evalúe el desarrollo de los modelos existentes, su riesgo de sesgo, su validación externa, así como su aplicabilidad e integración de los mismos en la práctica clínica.

En 2019 realizamos una revisión sistemática en MEDLINE, Scopus y EMBASE y seleccionamos artículos científicos que desarrollaran modelos pronósticos de mortalidad en pacientes con sepsis. De los artículos incluidos en la revisión extrajimos la información de los once dominios CHARMS y aplicamos la herramienta PROBAST, la cual valora tanto el riesgo de sesgo como la aplicabilidad de los modelos predictivos incluidos en una revisión sistemática. Realizamos un análisis descriptivo tanto de los dominios de CHARMS como de PROBAST. También revisamos si nuestros modelos habían sido validados externamente y comparamos nuestros resultados con estudios similares de otras patologías que utilizaban el checklist PROBAST.

De los 3.397 estudios cribados se escogieron 14 para su análisis. Los trabajos seleccionados mostraron una gran variabilidad en los 11 dominios CHARMS. Tras realizar la lista de verificación de PROBAST, observamos que solo un trabajo presentaba una valoración positiva respecto a la aplicabilidad junto con un riesgo de sesgo incierto, por no indicar cómo se manejaron los datos perdidos. El resto de los estudios presentaron un riesgo de sesgo alto. La valoración de la aplicabilidad fue satisfactoria en 6 de los 14 modelos, y la mayoría de ellos, integraban predictores que son de uso común en la práctica clínica habitual. Evaluamos la discriminación de los modelos, con un área bajo la curva *receiver operating characteristic* (ROC) entre 0,59 y 0,955, y no observamos ninguna falta de calibración. Solo tres modelos fueron validados externamente por otros autores, habiendo sido sometido uno de ellos a múltiples estudios de validación. En los estudios consultados sobre otras enfermedades, al igual que en el nuestro, dentro de la lista PROBAST, el dominio que presentaba un mayor riesgo de sesgo fue el de *Análisis*.

Hemos comprobado la importancia tanto de cumplir los estándares de calidad, como el rendimiento de los modelos y su validación externa. Podemos recomendar, aunque con cautela, la utilización de tres modelos, especialmente uno que ha sido sometido a múltiples estudios de validación externa. Con una validación externa adecuada, se podría considerar la implementación del resto de los modelos en la práctica clínica.

4.RESUMEN EN INGLÉS





Sepsis is an organ dysfunction secondary to an infection with high incidence, constituting one of the main causes of morbidity and mortality worldwide. Early identification and management are key to improving survival rates, especially in critically ill patients.

Stratification of the severity of sepsis using risk prediction models is essential to tailor treatment for better outcomes. There are many prognostic models available, but few are implemented in clinical practice. Their use is recommended, but since they will have a clinical impact, it is necessary that these models are developed according to international recommendations and previously validated in other populations. Furthermore, it is important to evaluate their applicability or potential use in routine clinical practice.

We aimed to conduct a systematic review of predictive mortality models in patients with sepsis, in accordance with international recommendations for such reviews, such as the CHARMS (Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies) and PROBAST (Prediction model Risk Of Bias Assessment Tool) guidelines, with the objective of obtaining a critical document that evaluates the development of existing models, their risk of bias, external validation, as well as their applicability and integration into clinical practice.

In 2019, we conducted a systematic review in MEDLINE, Scopus, and EMBASE and selected scientific articles that developed prognostic models for mortality in patients with sepsis. From the articles included in the review, we extracted information on the eleven CHARMS domains and applied the PROBAST tool, which assesses both the risk of bias and the applicability of predictive models included in a systematic review. We conducted a descriptive analysis of both CHARMS and PROBAST domains. We also checked if our models had been externally validated and compared our results with similar studies of other diseases that used the PROBAST checklist.

Out of the 3,397 screened studies, 14 were chosen for analysis. The selected studies showed great variability in the 11 CHARMS domains. After performing the PROBAST checklist, we observed that only one study had a positive assessment regarding applicability, but with an uncertain risk of bias since it did

not indicate how missing data were handled. The rest of the studies presented a high risk of bias. The assessment of applicability was satisfactory in 6 out of the 14 models, and most of them integrated predictors that are commonly used in routine clinical practice. We evaluated the discrimination of the models, with an area under the receiver operating characteristic (ROC) curve ranging from 0.59 to 0.955, and did not observe any lack of calibration. Only three models were externally validated by other authors; one of them had undergone multiple validation studies. In the consulted studies on other diseases, similar to ours, the domain with the highest risk of bias within the PROBAST checklist was the Analysis domain.

We have confirmed the importance of adhering to quality standards, as well as the performance of the models and their external validation. We can cautiously recommend the use of three models, especially one that has undergone multiple external validation studies. Implementation of the remaining models in clinical practice can be considered after appropriate external validation.



5.INTRODUCCIÓN





5.1. Sepsis

5.1.1. Antecedentes y epidemiología

Las enfermedades infecciosas conforman un importante problema de salud en el mundo siendo una de las principales causas de morbilidad y mortalidad.^{1,2} Se ha observado un aumento de la incidencia en las infecciones en los últimos años,³ debido probablemente a un aumento en la expectativa de vida, un aumento en la realización de técnicas invasivas y por estados de inmunosupresión secundarios a fármacos, factores que ocasionan un aumento de la susceptibilidad a la infección.¹

La sepsis es un trastorno orgánico potencialmente mortal, secundario a la infección, que afecta a millones de personas en el mundo cada año.^{4,5} La tasa de incidencia global de sepsis oscila entre 300 y 1.031 casos por 100.000 habitantes/año, con un incremento anual de un 13%.⁶ En España se ha constatado una tasa de incidencia anual de 367 casos de sepsis por 100.000 habitantes con una mortalidad del 12,8% de los mismos.⁷ Esta patología provoca un alto porcentaje de ingresos en las unidades de cuidados intensivos (UCI) y altos costes a los sistemas de salud, siendo una de las principales causas de mortalidad, la cual ocurre entre un tercio y un sexto de los afectados.^{4,5} Se estima que la sepsis provoca la muerte a decenas de millones de personas anualmente en todo el mundo.⁸

Además, existe un creciente conocimiento de que los pacientes que sobreviven a la sepsis tienen a menudo discapacidades psíquicas, psicológicas y cognitivas a largo plazo, con implicaciones sociales y de atención médica.⁹

Al igual que en otras patologías graves, como el politraumatismo, el infarto de miocardio o el ictus, la identificación y el manejo apropiado precoz de la sepsis puede mejorar la supervivencia.^{4,5,10}

5.1.2. Definición de sepsis y otros conceptos

A lo largo de las últimas décadas se han ido produciendo cambios en la definición de sepsis. En 1992 se publica el primer consenso del American College of Chest Physicians (ACCP)/Society of Critical Care Medicine (SCCM)¹¹, con el objetivo de estandarizar la terminología y definir los conceptos de sepsis y de fallo orgánico, recomendando el manejo de las siguientes definiciones:

- Infección: respuesta inflamatoria por la presencia de microorganismos o la invasión por estos microorganismos de un tejido del huésped normalmente estéril.
- Bacteriemia: presencia de bacterias viables en la sangre. La presencia de otros patógenos se define de la misma manera (viremia, fungemia, parasitemia).
- Septicemia, definida previamente como la presencia de bacterias o sus toxinas en la sangre. Se recomienda abandonar el uso de este término ya que puede llevar a confusión y no define adecuadamente el espectro completo de organismos patógenos que pueden infectar la sangre.
- Sepsis: es la respuesta inflamatoria sistémica asociada a una infección.
- Síndrome de respuesta inflamatoria sistémica (SRIS): es la presencia de dos o más de las siguientes: 1) temperatura mayor de 38°C o menor de 36°C; 2) frecuencia cardíaca mayor de 90 latidos por minuto; 3) frecuencia respiratoria (FR) mayor de 20 respiraciones por minuto, o hiperventilación manifestada por una presión parcial de dióxido de carbono ($p\text{CO}_2$) menor de 32mmHg; 4) recuento de leucocitos mayor de 12.000cel/ μl o menor de 4.000 cel/ μl ó más de 10% de formas inmaduras.
- Sepsis severa: cuando la sepsis se asocia a disfunción de órgano, hipoperfusión o hipotensión. La hipoperfusión se manifiesta por la presencia de acidosis láctica, oliguria o alteración aguda del estado mental. La hipotensión inducida

por sepsis se define como la tensión arterial sistólica (TAS) <90mmHg o un descenso > 40mmHg de la TAS basal en ausencia de otras causas de hipotensión.

- Shock séptico: sepsis con hipotensión a pesar de una adecuada resucitación con fluidos, junto con alteración de la perfusión (que puede incluir acidosis láctica, oliguria o alteración aguda del estado mental) o disfunción orgánica.

En el año 2001, las definiciones previas son revisadas por varias sociedades científicas de cuidados intensivos de Norteamérica y Europa: la SCCM, la ACCP, la European Society of Intensive Care Medicine (ESICM), la American Thoracic Society (ATS), y la Surgical Infection Society (SIS).¹⁰ En el documento de esta conferencia internacional reconocen que los criterios diagnósticos de SRIS publicados en 1992 son muy sensibles pero poco específicos. Ante la ausencia de evidencia científica para el cambio en las definiciones, y la ausencia de un criterio de referencia para el diagnóstico de sepsis, estas sociedades recomiendan mantener los conceptos de sepsis, sepsis severa y shock séptico. Aunque reconocen que estas definiciones no permiten estadiaje o pronóstico de la respuesta del huésped a la infección, no las modifican, pero aumentan la lista de criterios diagnósticos de sepsis con más parámetros generales, inflamatorios, hemodinámicos y parámetros de perfusión tisular (Tabla 1), indicando que, aunque estos hallazgos tampoco son específicos de sepsis, pueden ayudar a sospecharla.^{10,12}

Tabla 1. Criterios diagnósticos de sepsis 2001.^{10,12}

Infección, confirmada o sospechada, más algunos de los siguientes
<p>Parámetros generales</p> <ul style="list-style-type: none"> • Fiebre (temperatura > 38,3°C) • Hipotermia (temperatura < 36°C) • Frecuencia cardiaca > 90 latidos/minuto o 2 DE por encima del valor normal para la edad • Taquipnea > 30 respiraciones/minuto • Alteración del estado mental • Edema significativo o balance hídrico positivo (> 20ml/kg en 24 horas) • Hiperglucemia > 110mg/dl (en ausencia de diabetes)
<p>Parámetros inflamatorios</p> <ul style="list-style-type: none"> • Leucocitosis (recuento de glóbulos blancos > 12.000/μl) • Leucopenia (recuento de glóbulos blancos < 4.000/μl) • Recuento de glóbulos blancos normal con > 10% de formas inmaduras • Proteína C reactiva en plasma > 2 DE sobre el valor normal • Procalcitonina en plasma > 2 DE sobre el valor normal
<p>Parámetros hemodinámicos</p> <ul style="list-style-type: none"> • Hipotensión arterial (TAS < 90mmHg, TAM < 70mmHg o disminución de TAS > 40mmHg en adultos o un valor menor de 2 DE por debajo de lo normal para la edad) • Saturación venosa mixta de oxígeno > 70% • Índice cardiaco > 3,5L/min/m²
<p>Parámetros de disfunción orgánica</p> <ul style="list-style-type: none"> • Hipoxemia arterial (PaO₂/FiO₂ < 300) • Oliguria aguda (diuresis < 0,5ml/kg/hora durante al menos 2 horas) • Incremento de la creatinina > 0,5mg/dl • Alteración de la coagulación (INR > 1,5 o TPTA > 60 segundos) • Íleo (ausencia de ruidos hidroaéreos) • Trombopenia (plaquetas < 100.000 μl) • Hiperbilirrubinemia (bilirrubina total en plasma > 4mg/dl)
<p>Parámetros de perfusión tisular</p> <ul style="list-style-type: none"> • Hiperlactatemia • Disminución del relleno capilar o moteado cutáneo

Abreviaturas: DE, desviación estándar; FiO₂, fracción de oxígeno inspirado; INR, international normalized ratio; PaO₂, presión arterial de oxígeno; TAM, tensión arterial media; TAS, tensión arterial sistólica; TPTA, tiempo parcial de tromboplastina activada.

Tras la segunda conferencia de 2001, al ampliar la lista de posibles criterios diagnósticos, se corría el riesgo de que la definición fuera todavía menos específica.¹³

Todo ello originaba una baja especificidad de los criterios diagnósticos de sepsis, ya que una respuesta inflamatoria sistémica podía aparecer en otras patologías no infecciosas, tales como infarto agudo de miocardio, trauma severo, pancreatitis o grandes quemados.¹³⁻¹⁵

Puesto que la sepsis se definía por la presencia de criterios SRIS más una infección sospechada o documentada, y dado que los pacientes con otras enfermedades agudas no infecciosas cumplen los criterios SRIS, podemos pensar que la sepsis equivale a infección. Pero, aunque todos los pacientes con sepsis tienen una infección, no todos los pacientes con infección tienen sepsis, ni la van a desarrollar. Además, no está claro el papel de la infección en la patogenia del SRIS, pues como hemos visto, una inflamación estéril como la producida en un trauma severo, en un gran quemado o en una pancreatitis, pueden provocar signos clínicos de inflamación sistémica, porque la respuesta inmune innata y adaptativa aparece tanto en respuesta a la infección como en la reparación de tejido lesionado sin infección.¹³

Es más, cualquier infección se asocia habitualmente con fiebre, taquicardia, leve hiperventilación y leucocitosis, y su ausencia puede sugerir una inmunodepresión del huésped y no necesariamente ausencia de sepsis. Así, una persona inmunodeprimida puede presentar sepsis y no desarrollar esos signos típicos de infección.¹³

La sepsis se puede considerar como una respuesta nociva del huésped a la infección, que origina una disfunción orgánica.¹³ Ya en 2010, en el Simposio Merinoff (reunión internacional sobre sepsis), definían la sepsis como “una enfermedad potencialmente mortal que se produce cuando la respuesta del huésped a la infección daña sus propios órganos y tejidos”. Se hace referencia a una definición de la sepsis desde el punto de vista molecular en la que, la presencia de moléculas del huésped junto con los productos de la infección provocaría una respuesta alterada de la inmunidad con liberación de mediadores inmunes que originarían los síntomas y signos de la sepsis.⁸

Se pone en entredicho, que la sepsis sea tan solo, una respuesta inflamatoria con infección, siendo más bien un síndrome de difícil definición, caracterizado por los diferentes procesos fisiopatológicos e inmunológicos involucrados, muchos de ellos todavía desconocidos y cuya presentación clínica varía dependiendo de otros factores como la edad del paciente, su comorbilidad, estado inmunológico y tratamiento.¹²

Es necesario mejorar y consensuar unos criterios diagnósticos para identificar con precisión a los pacientes con sepsis,¹⁵ y facilitar la investigación futura sobre sepsis, con el fin de disponer de criterios claros y precisos para seleccionar a estos pacientes en los diferentes estudios o ensayos clínicos (p.ej. investigación de nuevas terapias).¹⁴

Tras más de dos décadas sin prácticamente cambios en las definiciones de sepsis y shock séptico, en 2016 se publica la actualización de estas definiciones en el tercer documento de consenso internacional, conocido como Sepsis-3. En este documento un grupo de expertos de la SCCM Y ESICM desarrollan las nuevas definiciones de sepsis y shock séptico que fueron revisadas y respaldadas por otras 31 sociedades científicas internacionales.³

Este grupo de expertos reconoce que no existe una prueba diagnóstica validada para sepsis (“gold standard”), y que la fisiopatología de la sepsis no está totalmente aclarada. Se considera que la disfunción orgánica se produce por la respuesta del huésped a la infección. Los criterios SRIS previos no indican necesariamente una respuesta anómala del huésped, sino que pueden reflejar una respuesta apropiada del huésped a la infección. Incluso estos criterios SRIS, como ya hemos comentado, aparecen en pacientes hospitalizados que no desarrollan infección, y no identifican a todos los pacientes ingresados en UCIs con infección más fallo orgánico, a pesar de tener estos pacientes un curso clínico con elevada mortalidad y morbilidad.³

Por todo ello, se siguen buscando nuevos criterios clínicos y definiciones que sean claros, útiles y válidos, que puedan identificar todos los elementos de la sepsis (la infección, la respuesta del huésped y la disfunción orgánica), que sean fáciles de obtener, disponibles de forma precoz y a un bajo coste.

Además, se debería disponer de criterios clínicos que identificaran precozmente a pacientes con sospecha de infección que pudieran tener o desarrollar una sepsis, para poder ser utilizados en todos los niveles asistenciales tanto en entornos hospitalarios como extrahospitalarios. Mientras tanto, los criterios inespecíficos del SRIS como fiebre o neutrofilia, continuarán ayudando en el diagnóstico general de la infección y se complementarán con características de infecciones específicas que nos orientarán para localizar el foco de la infección o el microorganismo causante.³

Para medir la disfunción orgánica, existen diferentes sistemas de puntos, que consideran tanto los hallazgos clínicos, de laboratorio o, las medidas terapéuticas adoptadas. El Sequential Organ Failure Assessment (SOFA) es uno de los sistemas de puntuación más extendido para medir la disfunción orgánica, por su fácil manejo. La disfunción orgánica se define por un aumento de 2 o más puntos en la puntuación de la escala SOFA (Tabla 2), circunstancia que se relaciona con una mortalidad hospitalaria superior al 10%.³

De esta forma, los criterios clínicos de sepsis quedan definidos por padecer una infección, sospechada o confirmada, junto con un aumento en dos o más puntos en la escala SOFA basal.

Tabla 2. Escala de medición de disfunción orgánica: *Sequential Organ Failure Assessment (SOFA)*.^{3,16}

Sistema	Puntuación				
	0	1	2	3	4
Respiración PaO ₂ /FiO ₂ (mmHg)	≥400	<400	<300	<200 Con soporte respiratorio	<100 Con soporte respiratorio
Coagulación Plaquetas (10 ³ /mm ³)	≥150	<150	<100	<50	<20
Hígado Bilirrubina (mg/dl)	<1,2	1,2-1,9	2,0-5,9	6,0-11,9	>12,0
Cardiovascular TAM (mmHg) o su manejo	TAM≥70	TAM<70	Dopamina≤5 o dobutamina (cualquier dosis)*	Dopamina>5 o epinefrina≤0,1 o norepinefrina≤0,1*	Dopamina>15 o epinefrina>0,1 o norepinefrina>0,1*
SNC Escala de Glasgow	15	13-14	10-12	6-9	<6
Renal Creatinina (mg/dl) o Flujo urinario (mL/día)	<1,2	1,2-1,9	2,0-3,4	3,5-4,9 <500	>5,0 <200

Abreviaturas: FiO₂, fracción de oxígeno inspirado; PaO₂, presión arterial de oxígeno; SNC, sistema nervioso central; TAM, tensión arterial media.

*Medicamentos adrenérgicos en dosis de µg/Kg/min, administrados durante al menos una hora.

Algunos componentes de la escala SOFA requieren pruebas de laboratorio, por lo que se necesita un tiempo para su aplicación hasta obtener todos estos datos. Son pruebas que generalmente se realizan de rutina en la atención inicial hospitalaria de estos pacientes.

El grupo de expertos del documento Sepsis-3 dejaron de recomendar el uso del término *sepsis severa*, por ser redundante, y actualizaron la definición del shock séptico como un subgrupo de la sepsis en el que las alteraciones circulatorias, metabólicas y celulares son más marcadas y se asocia con un mayor riesgo de mortalidad. Clínicamente se identifica a los pacientes con shock séptico porque, a pesar de una adecuada resucitación con fluidos, es decir, en ausencia de hipovolemia, precisan vasopresores para mantener una tensión arterial media (TAM) igual o mayor de 65 mmHg y presentan unos niveles de lactato sérico por encima de 2 mmol/L (>18 mg/dL) en relación con la disfunción celular y el compromiso cardiovascular, e identifican mejor el shock séptico que cualquiera de ellos por separado. La presencia de estos dos parámetros, TAM y niveles de lactato, se asocia con una tasa de mortalidad hospitalaria superior al 40%.³

Estas nuevas definiciones, además de las implicaciones clínicas, tienen implicaciones epidemiológicas y de investigación, pudiendo adoptarse como criterios de inclusión en ensayos clínicos. La estandarización de las definiciones y de los criterios clínicos es crucial para conseguir una comunicación científica clara y una mejor identificación de esta patología.³

5.1.3. Recomendaciones en el tratamiento de la sepsis

Desde 2004 se han publicado cinco ediciones de las guías internacionales “Campaña Sobrevivir a la Sepsis” (CSS), con recomendaciones para el tratamiento de la sepsis y el shock séptico. El objetivo de estas directrices es servir como guía para los médicos que tratan a pacientes con esta patología, para mejorar sus resultados.

La última edición se publica en 2021, con la evidencia científica existente en la literatura hasta julio de 2019, y es elaborada por un comité de 60 expertos con

el respaldo de 23 sociedades internacionales.⁵ En esta edición se realiza un total de 93 recomendaciones (fuertes o débiles, en base a su evidencia científica) sobre diferentes aspectos en el tratamiento de la sepsis, como son la detección y reanimación inicial, el tratamiento de la infección, el manejo hemodinámico, la ventilación, las terapias adicionales y los objetivos de atención y los resultados a largo plazo. Algunas de las recomendaciones son declaraciones de buenas prácticas cuando el beneficio o el daño son inequívocos, pero no existe o es difícil evaluar la evidencia científica.⁵

Los expertos consideran la sepsis y el shock séptico como emergencias médicas y recomiendan iniciar el tratamiento y la reanimación de forma precoz. Estas guías recomiendan la existencia de programas de mejora en sepsis en los hospitales, que incluyan la detección precoz de sepsis en pacientes con enfermedades agudas, la detección de pacientes de alto riesgo y el uso de procedimientos o paquetes de tratamiento estándar.^{4,5}

5.1.3.1 Detección precoz de sepsis

El reconocimiento precoz de la enfermedad resulta imprescindible para iniciar el conjunto de medidas terapéuticas incluidas en la reanimación inicial, idealmente, dentro de la primera hora tras la sospecha.^{8,17}

Este diagnóstico precoz de la sepsis es especialmente importante en los pacientes más graves.¹⁸

Para la sospecha e identificación precoz de pacientes con mayor riesgo de sepsis existen diferentes variables clínicas y herramientas que se pueden aplicar fácilmente a pie de cama¹⁸ como pueden ser los criterios SRIS, signos vitales, signos de infección y la puntuación en diferentes escalas como, la escala quick SOFA (qSOFA) (tabla 3), y las escalas de alerta temprana como son la National Early Warning Score (NEWS) (tabla 4) y la Modified Early Warning Score (MEWS) (tabla 5). Cada herramienta con una sensibilidad y especificidad diferente para la detección precoz de sepsis.⁵

Como método inicial de screening nos va a ser útil una escala con datos clínicos que se puedan obtener de forma rápida en la valoración inicial a la cabecera del paciente. En el tercer documento de consenso internacional de sepsis se recomienda el uso del qSOFA tanto como predictor de una mala evolución en pacientes con infección, sospechada o confirmada, como en la valoración inicial del paciente; permite su uso a pie de cama de forma rápida y precoz, alertando a los médicos sobre la presencia de infección en pacientes sin sospecha inicial, presencia de disfunción orgánica, necesidad de un tratamiento precoz o más intensivo y para derivar al paciente a una unidad adecuada para su tratamiento y monitorización.³

La escala qSOFA es una versión simplificada de la escala SOFA que discrimina mortalidad hospitalaria, ligeramente mejor que la escala SOFA o el SRIS, en pacientes con sospecha de infección fuera de la UCI.^{3,19} Esta escala consta de tres criterios clínicos: un aumento de la frecuencia respiratoria, una alteración del estado mental y una disminución de la TAS. Los pacientes con un peor pronóstico, o con una posible peor evolución, serían aquellos con sospecha de infección que cumplan al menos dos de los tres criterios de esta escala.³

Tabla 3. Escala qSOFA.³

Frecuencia respiratoria ≥ 22
Alteración estado mental: Glasgow < 15
Tensión arterial sistólica ≤ 100

qSOFA: quick Sequential Organ Failure Assessment

Numerosos estudios han investigado el uso de qSOFA como herramienta de detección de sepsis con resultados contradictorios, viendo que es más específica pero menos sensible que tener dos de los cuatro criterios SRIS para la

identificación precoz de disfunción orgánica debida a infección. Hallazgos parecidos se obtuvieron al compararlo con las escalas NEWS y MEWS.⁵ Estas dos escalas, NEWS y MEWS, también utilizan datos clínicos de fácil obtención a la cabecera del paciente, pero tienen un mayor valor predictivo de deterioro clínico e ingreso en UCI que la escala qSOFA en pacientes con sospecha de infección fuera de la UCI. La escala NEWS ha demostrado una mayor precisión, por lo que la escala qSOFA no debería reemplazar a estas escalas en la valoración inicial del riesgo de pacientes con sospecha de infección.²⁰

Recientemente, en la actualización publicada en 2021 de la CSS, incluyen una nueva recomendación respecto a las ediciones previas, en la que desaconsejan el uso de la escala qSOFA en comparación con SRIS, NEWS o MEWS, como única herramienta de detección de sepsis o shock séptico. Aunque un qSOFA positivo nos debe alertar sobre la posibilidad de sepsis en los diferentes entornos de la atención sanitaria, no recomiendan utilizar esta escala como única herramienta de detección de sepsis por su baja sensibilidad.⁵



Tabla 4 National Early Warning Score (NEWS)²¹

Parámetro fisiológico	3	2	1	0	1	2	3
Frecuencia respiratoria	≤ 8		9 - 11	12 - 20		21 - 24	≥ 25
SpO2 (%)	≤ 91	92 - 93	94 - 95	≥ 96			
Oxígeno suplementario		Yes		No			
Temperatura (°C)	≤ 35		35,1 – 36,0	36,1 – 38,0	38,1 – 39,0	≥ 39,1	
TAS (mmHg)	≤ 90	91 - 100	101 – 110	111 - 219			≥ 220
Frecuencia cardíaca	≤ 40		41 - 50	51 - 90	91 - 110	111 - 130	≥ 131
Nivel de consciencia				Alerta			V, P o U

Abreviaturas: SpO2, saturación de oxígeno; TAS, tensión arterial sistólica; V, responde a estímulo verbal; P, responde al dolor; U, no responde.

Tabla 5 Modified Early Warning Score (MEWS)²²

Parámetro fisiológico	3	2	1	0	1	2	3
TAS (mm Hg)	< 70	71-80	81-100	101 - 199		≥ 200	
Frecuencia cardíaca		< 40	41 - 50	51 - 100	101 - 110	111 - 129	≥ 130
Frecuencia respiratoria		< 9		9 - 14	15 - 20	21 - 29	≥ 30
Temperatura (°C)		< 35		35 – 38,4		≥ 38,5	
Escala AVPU				Alerta	V	P	U

Abreviaturas: TAS, tensión arterial sistólica; Escala AVPU, valoración nivel de consciencia; V, responde a estímulo verbal; P, responde al dolor; U, no responde.

De las escalas mencionadas, la escala NEWS actualmente es la herramienta estándar de alerta temprana para la activación de los equipos de respuesta rápida, en la atención de cualquier patología médica aguda, en los hospitales de Reino Unido. Indican una puntuación de 7 como punto de corte para poner en alerta sobre una situación clínica de riesgo alto de complicación o mortalidad, y una puntuación global de 5 o de 3 en uno solo de los parámetros para alertar sobre una situación de riesgo medio.²⁰

Respecto a la escala MEWS establecen que una puntuación de 5 o superior, en enfermos con patología urgente, conlleva un aumento del riesgo de mortalidad o ingreso en UCI.²²

A pesar de la variabilidad en la sensibilidad y especificidad de las diferentes escalas para sospecha de sepsis, no dejan de ser una herramienta importante para identificar sepsis precozmente e instaurar el tratamiento oportuno.⁵

El estándar actual, en nuestro medio, para el reconocimiento precoz de sepsis es el uso de la escala qSOFA, y la escala SOFA para el diagnóstico de sepsis y la valoración de la disfunción orgánica.^{23,24} Cada hospital puede emplear además otros sistemas que considere convenientes para la detección precoz de infección, como el SRIS, y para la valoración de deterioro clínico de cualquier causa, como es la puntuación NEWS.^{24,25}

Al igual que en otras enfermedades tiempo dependientes, se ha desarrollado el concepto de *Código Sepsis*, con el objetivo de identificar precozmente al enfermo con sepsis en todos los ámbitos asistenciales y hacer un abordaje multidisciplinar que garantice la rápida implantación de las medidas que han demostrado reducir la mortalidad por esta enfermedad.²⁴⁻²⁶ En el año 2014 se desarrolla un documento de consenso, con la colaboración de 15 sociedades científicas españolas, cuya finalidad era favorecer la implantación y mejora de un código sepsis en los sistemas sanitarios nuestro país.^{23,27}

En los últimos años, se ha implantado el código sepsis en más del 50% de hospitales en España, utilizando en la mayoría del ellos el qSOFA como criterio

para su activación, la cual puede ser llevada a cabo en los diferentes niveles asistenciales, tanto por personal de enfermería como por personal médico.²³

Existen datos del impacto positivo que genera la implantación del código sepsis en los servicios sanitarios, con una mejora significativa en el cumplimiento de las recomendaciones internacionales de la campaña para sobrevivir a la sepsis y con un descenso de la mortalidad a corto plazo.^{28,29}

5.1.3.2. Tratamiento de la sepsis

Además de las guías de práctica clínica, en 2018, los expertos de la CSS desarrollaron un paquete de medidas a aplicar en la primera hora tras el reconocimiento de la sepsis. Este conjunto de procedimientos se debe implementar con el objetivo de mejorar la atención y los resultados en estos pacientes.³⁰

El paquete de medidas de la primera hora incluye la medición del nivel de lactato, la obtención de hemocultivos previo al inicio del tratamiento antibiótico, administración de antibióticos de amplio espectro, la infusión rápida de 30 ml/kg de cristaloides ante la presencia de hipotensión o lactato elevado y la administración de aminas vasoactivas si con la fluidoterapia no se consigue una TAM mayor o igual de 65mmHg.^{5,30}

En lo referente al uso de paquetes estándar de tratamiento de la sepsis, el documento de 2021 de la CSS recomienda la aplicación de medidas precoces que incluyen la identificación temprana, la medición de los niveles de lactato en sangre, extracción de cultivos, inicio de antibioticoterapia y el tratamiento con fluidos. Estos últimos son componentes del paquete de medidas para la primera hora. Recomiendan la medición del nivel de lactato en sangre cuando se sospeche sepsis, ya que su asociación con la mortalidad está bien establecida, y es parte de la definición del shock séptico.⁵

Respecto al inicio del tratamiento antibiótico, en pacientes con posible shock séptico o una alta probabilidad de sepsis, se recomienda su administración

precoz, en el plazo máximo de una hora.⁵ Es conocido que el retraso de la antibioterapia en la sepsis se asocia a una mayor mortalidad hospitalaria.¹⁷

En cuanto al tratamiento inicial con fluidos recomiendan la administración intravenosa de al menos 30 ml/kg de cristaloides dentro de las tres primeras horas de reanimación en los pacientes que presenten hipoperfusión o shock séptico. Sugieren guiar esta reanimación con líquidos con la valoración de parámetros dinámicos (la respuesta del volumen sistólico y su variación, los cambios en la presión de pulso o la ecocardiografía tras maniobras como la elevación pasiva de las piernas o la administración de un bolo rápido de fluidos), en lugar de guiarnos solo por el examen físico o parámetros estáticos (frecuencia cardíaca, TAS y presión venosa central). Otros parámetros que pueden guiar la reanimación inicial serían la disminución del lactato sérico en pacientes con sepsis o shock séptico con niveles iniciales elevados, o el tiempo de relleno capilar en pacientes con shock séptico.⁵

Estas guías recogen también recomendaciones sobre la elección del tratamiento antibiótico empírico inicial, así como su desescalada y duración; indicaciones sobre el control urgente del foco; el manejo hemodinámico con recomendaciones sobre el uso de fluidos y de agentes vasopresores e inotrópicos; ventilación mecánica; terapias adicionales como corticoides, hemoderivados, bicarbonato, profilaxis tromboembólica y gastroduodenal, tratamiento renal sustitutivo, control de glucemia y nutrición. Además realizan recomendaciones sobre los cuidados paliativos en pacientes con sepsis, apoyo y educación para paciente y familias, terapia cognitiva y seguimiento al alta.⁵

Hay pruebas contundentes que han demostrado que los paquetes de reanimación intensiva, el control adecuado del foco de infección, el tratamiento antibiótico adecuado y el soporte orgánico son la piedra angular del éxito en el tratamiento de los pacientes con sepsis.¹⁷

La estratificación de la gravedad de la sepsis mediante modelos de riesgo es primordial para adecuar el tratamiento con el fin de obtener unos mejores resultados.^{4,17,18}

5.1.4. Factores de riesgo de infección, sepsis, fallo orgánico y mortalidad.

Un factor pronóstico es cualquier variable o medida que, entre las personas con una condición de salud determinada, se asocia con un resultado clínico posterior.³¹ Existen multitud de factores de riesgo independientes relacionados con las infecciones (Tabla 6).

Tabla 6

Factores de riesgo relacionados con las infecciones
Relacionados con las infecciones más prevalentes^{1,2}
Respiratoria: varones, edad media elevada, broncopatías, cardiopatías, diabetes mellitus, tabaquismo, adicción a drogas vía parenteral, tratamiento inmunosupresor. Urinaria: mujeres, diabetes, portador de sonda vesical, nefro-uropatías, portador de catéter, inmunodeficiencias.
Relacionados con patógenos multirresistentes^{1,4}
Ingreso hospitalario previo en los últimos 3 meses. Uso de antibiótico en el último mes. Procedencia de una institución de cuidados crónicos o residencia de ancianos. Colonización o infección previas por organismos multirresistentes. Tratamiento inmunosupresor. Portador de sonda vesical o de catéter central.
Relacionados con la mortalidad¹⁸
Enfermedad terminal. Cáncer metastásico. Hipotensión arterial. Dificultad respiratoria (taquipnea o hipoxemia). Recuento bajo de plaquetas, bandemia (>5% de neutrófilos inmaduros o cayados). Edad mayor de 65 años. Sospecha de una infección de las vías respiratorias inferiores. Residencia en un hogar de ancianos y estado mental alterado. Variabilidad de la frecuencia cardíaca.³² Retraso en el inicio del tratamiento antibiótico. Disfunción orgánica.³³

La localización de la infección también predispone al desarrollo de sepsis. Por orden de riesgo de desarrollar sepsis se sitúan en primer lugar las infecciones neurológicas, seguidas de la neumonía adquirida en la comunidad, la pielonefritis y la infección intraabdominal.^{1,2}

El análisis de la variabilidad de la frecuencia cardiaca como factor de riesgo de mortalidad en sepsis se basa en que la presencia de una variabilidad sustancial de la frecuencia cardiaca, parece reflejar una interacción saludable entre los diversos moduladores del cuerpo, como el nodo sinusal y el centro respiratorio; con lo que un descenso en la variabilidad de la frecuencia cardiaca puede predecir mayor gravedad y mortalidad en sepsis.³² En relación con el retraso en el tratamiento antibiótico hay un aumento, estadísticamente significativo, en la probabilidad de muerte asociado con el número de horas de retraso de la primera administración de antibiótico, y estos resultados fueron similares en pacientes con sepsis grave y shock séptico.^{17,19}

La presencia de disfunción orgánica indica que estamos ante un proceso potencialmente mortal.¹³ En una infección, un incremento en la puntuación de las escalas de disfunción orgánica a lo largo del tiempo, se correlacionan con una mayor mortalidad aunque no sean los idóneos, ya que para predecir mortalidad se deben usar los scores desarrollados con tal fin.³³

Existen varios modelos de riesgo que evalúan la presencia y severidad de disfunción orgánica, entre ellos destacan tres por ser los más usados en población general en UCI: escala SOFA, Logistic Organ Dysfunction System (LODS) y Multiple Organ Dysfunction Score (MODS). Todos ellos incluyen disfunción de seis órganos: cardiovascular, respiratorio, renal, neurológico, hepático y hematológico (Tabla 7):³³

Tabla 7³³

Factores relacionados con disfunción orgánica
Cardiovascular
FC, FC ajustada a la tensión arterial TAS, TAM Uso de vasopresores
Respiratorio
PaO ₂ / FiO ₂ Ventilación mecánica
Renal
Creatinina Urea o BUN Diuresis
Neurológico
Escala de coma de Glasgow
Hepático
Bilirrubina Tiempo de protrombina
Hematológico
Recuento de plaquetas Recuento de leucocitos

Abreviaturas: FC, frecuencia cardiaca; TAS, tensión arterial sistólica; TAM, tensión arterial media; PaO₂, presión arterial de oxígeno; FiO₂, fracción de oxígeno inspirado; BUN, nitrógeno ureico.

Las escalas que evalúan la gravedad de la enfermedad, que son útiles para predecir mortalidad o supervivencia en pacientes críticos, y que más se utilizan en la práctica clínica habitual son Acute Physiology and Chronic Health Evaluation (APACHE), Simplified Acute Physiology Score (SAPS) y Mortality Probability Model (MPM).³³ Las variables utilizadas en estos tres modelos son condiciones previas al ingreso, comorbilidades, variables del ingreso actual tanto clínicas como analíticas y tratamiento actual (Tabla 8).³⁴⁻³⁶

Tabla 8³³⁻³⁶

ESCALAS DE GRAVEDAD EN PACIENTES CRÍTICOS		
APACHE IV	SAPS 3	MPM-III
142 variables	20 variables	16 variables
Edad	Edad	Edad
Frecuencia cardíaca	Frecuencia cardíaca	Frecuencia cardíaca
TAM	TAS	TAS
Ventilación mecánica	Soporte ventilatorio/O ₂	Ventilación mecánica
Escala Glasgow	Escala Glasgow	Coma/estupor (GCS 3-4)
Creatinina y BUN	Creatinina	IRA
Diuresis	IRC	IRC
Insuficiencia hepática Tumor maligno, SIDA	Cirrosis Tumor maligno, SIDA	Cirrosis Neoplasia metastásica
Cirugía urgente	Patología actual, cirugía, infección nosocomial	Arritmia actual, accidente cerebrovascular, efecto masa intracraneal
Bilirrubina	Bilirrubina	
Temperatura	Temperatura	
pH	Acidosis	
Frecuencia respiratoria		
Oxigenación	Oxigenación	
	Uso fármacos vasoactivos	RCP previo al ingreso en UCI
Hematocrito	Trombopenia	Hemorragia GI
Leucocitos	Leucocitos	
Sodio, albúmina, glucosa		
Diagnóstico al ingreso	Motivo de ingreso en UCI	
Origen del paciente	Origen del paciente	
Duración estancia previa a UCI	Duración estancia previa a UCI	
Comorbilidades	Comorbilidades	Ausencia de otros factores de riesgo

Abreviaturas: APACHE, Acute Physiology and Chronic Health Evaluation; SAPS, Simplified Acute Physiology Score; MPM, Mortality Probability Model; TAM, tensión arterial media; TAS, tensión arterial sistólica; O₂, oxígeno; GCS, escala de Glasgow; BUN, nitrógeno ureico; IRA, insuficiencia renal aguda; IRC, insuficiencia renal crónica; RCP, reanimación cardiopulmonar; UCI, unidad de cuidados intensivos; GI, gastrointestinal.

Estos tres modelos que predicen mortalidad, al igual que los tres anteriores que predicen disfunción orgánica, se desarrollaron para su uso en grupos mixtos de pacientes de UCI y su precisión en subgrupos de pacientes, como podría ser una patología concreta, puede cuestionarse. Se están desarrollando cada vez más sistemas de puntuación específicos de diferentes enfermedades.³³

Como vemos, hay multitud de posibles variables que pueden pronosticar mortalidad en sepsis como: variables sociodemográficas (edad...), factores predisponentes del paciente (morbilidad previa como diabetes, cáncer, cáncer hematológico, tratamiento oncológico, inmunosupresión, cirrosis, SIDA...), constantes clínicas (tensión arterial, FC, FR, temperatura, escala de Glasgow, saturación o PaO₂, PaO₂/FiO₂...), características de la infección (patógenos, lugar de infección...), datos de laboratorio (lactato, pH, creatinina, leucocitos, plaquetas, coagulación, bilirrubina...), incluso tratamientos (antibióticos, necesidad de fluidoterapia, drogas vasoactivas o ventilación mecánica, cirugía urgente).^{1,2,4,16-19,32-36}

Dada la variabilidad entre pacientes, junto con su comorbilidad, y en la etiología, presentación y tratamiento de enfermedades, un solo predictor o variable rara vez proporciona una estimación adecuada del pronóstico. En la práctica clínica se utilizan múltiples predictores para estimar el pronóstico de un paciente.^{31,37}

Por tanto, es importante identificar qué combinación de todas estas variables determinan de forma segura e inequívoca la probabilidad de mortalidad en una patología determinada. Nosotros evaluaremos los diferentes modelos pronósticos en sepsis desarrollados para determinar qué modelo puede resultar de utilidad en la práctica clínica habitual, para adecuar el tratamiento y mejorar los resultados en estos pacientes.

5.2. Modelos predictivos.

Comprender y mejorar el pronóstico de una enfermedad o condición de salud es una prioridad en la investigación clínica y la práctica.³⁸ La investigación fundamental pronóstica es la investigación que tiene como objetivo examinar los

resultados de una enfermedad o condición de salud en el contexto de la práctica clínica actual y es un camino hacia la mejora de los resultados de los pacientes.³⁸

Hemos visto que categorizar la gravedad de los pacientes sépticos es un pilar fundamental para la identificación y el manejo estructurado precoz, sobre todo en aquellos pacientes con mayor riesgo de mortalidad.

Existen herramientas para conseguir este propósito, como son los modelos de predicción de riesgo o escalas pronósticas, en los que los profesionales médicos nos apoyamos para tomar decisiones clínicas.³⁸

5.2.1. Definición.

Un modelo pronóstico es una combinación formal de múltiples predictores a partir de la cual se pueden calcular los riesgos de un resultado final específico para pacientes individuales,³⁹ ayudándonos en la toma de decisiones con el objeto de mejorar los resultados clínicos en nuestros pacientes.^{31,39,40}

5.2.2. Utilidades y limitaciones.

Existen diversas utilidades de los modelos de predicción o pronósticos, tanto en el ámbito clínico, como en investigación y en gestión sanitaria.

Dentro del ámbito clínico los modelos predictivos nos ayudan a mejorar el conocimiento sobre el curso de la enfermedad y sus determinantes, orientan al médico en la toma de decisiones en cuanto al tratamiento y seguimiento de los pacientes con diferente riesgo pronóstico y facilitan una información más detallada y fiable a los pacientes sobre su pronóstico o sobre la gravedad de su enfermedad.³⁷

En el terreno de la investigación clínica los modelos pronósticos ayudan a mejorar el diseño y el análisis de ensayos clínicos aleatorizados.³⁹ Nos van a permitir estratificar el riesgo de los sujetos participantes en un ensayo clínico.⁴¹ Además, el poder estratificar o categorizar a los pacientes que sufren una

determinada enfermedad según su riesgo pronóstico permite realizar una selección más precisa de pacientes con un determinado riesgo para una investigación concreta (por ejemplo, seleccionar pacientes con mayor riesgo de mortalidad para ensayar determinado fármaco o tratamiento).³⁷ Incluso los resultados de los ensayos terapéuticos aleatorizados se pueden utilizar para estimar cómo un tratamiento específico modificaría el pronóstico estimado de un paciente con un determinado riesgo pronóstico.³⁹

Estos modelos también pueden ser útiles en el ámbito de la gestión sanitaria ya que nos van a permitir comparar resultados entre diferentes hospitales o servicios sanitarios.³⁷ Si bien los datos brutos de mortalidad pueden ofrecer alguna orientación global sobre el desempeño de un servicio en concreto, ajustar las tasas de mortalidad según la gravedad de la enfermedad, mediante el uso de puntuaciones de predicción de mortalidad, para calcular la tasa de mortalidad estandarizada, puede ayudar a mejorar la evaluación de la calidad. Estos indicadores ajustados a la gravedad se pueden utilizar para evaluar el rendimiento de un servicio médico o quirúrgico, o de un hospital, a lo largo del tiempo o para comparar varias unidades, hospitales o áreas de salud. Existen grandes variaciones en las tasas de mortalidad ajustadas al riesgo entre los hospitales, y la evaluación repetida de la calidad puede ayudar a determinar las razones subyacentes a estas diferencias y permitir el desarrollo de programas para mejorar el rendimiento.³³

Observamos cómo en los últimos años los modelos de predicción de riesgo en medicina han proliferado, y cada vez se recomienda más su inclusión y uso en las guías de práctica clínica.⁴¹

A pesar de la existencia de múltiples modelos pronósticos y de su potencial utilidad, pocos se utilizan de forma rutinaria en la práctica clínica,³⁹ no estando claro por qué los médicos son reacios a utilizarlos o por qué utilizan un determinado modelo y no otro.⁴² Las razones para el uso limitado de estos modelos podrían ser:

- Complejos, de difícil aplicación, que requieran mucho tiempo en su cumplimentación y que contengan variables difíciles de obtener en la práctica clínica habitual.³⁷

- Que existan dudas sobre el rendimiento del modelo por no cumplir unos mínimos de metodología en su desarrollo.⁴³
- Que no hayan sido validados en otras poblaciones.⁴³
- O que generen dudas sobre cómo utilizar las probabilidades predichas en la toma de decisiones por parte de los clínicos.³⁷

Si un modelo tiene deficiencias en su desarrollo o validación, sus estimaciones podrían ser imprecisas y, por tanto, ser la causa de la desconfianza en su uso.^{40,42} Los modelos que aportan más confianza para la práctica clínica son los que se desarrollan sobre una base de datos grande y de alta calidad, los que se basan en métodos estadísticos sólidos, y los que son validados con conjuntos de datos independientes obtenidos de diferentes localizaciones. Con frecuencia, los modelos publicados se han desarrollado utilizando métodos estadísticos poco adecuados.³⁹

Hay recomendaciones en la literatura de no utilizar un modelo pronóstico en la práctica clínica hasta que, como mínimo, su rendimiento haya sido sometido a validación externa.^{37,39}

Un modelo pronóstico puede influir en el resultado del paciente si se realizan cambios en el manejo clínico, en función de la información pronóstica proporcionada, condicionando un impacto clínico positivo o negativo.^{39,41} La evaluación del impacto clínico del uso de un modelo pronóstico requiere de estudios comparativos de cohortes en los que se compare un grupo al que se le proporciona los cuidados habituales sin el uso del modelo y otro grupo en el que las predicciones del modelo guíen las decisiones de tratamiento. Existen pocos estudios en la literatura que evalúen el impacto de un modelo pronóstico en la práctica clínica.³⁹

5.2.3. Calidad de los modelos pronósticos.

El desarrollo de un modelo es un proceso complejo, en el que se debe disponer de los detalles suficientes en cuanto a los métodos de desarrollo del modelo y al manejo de los datos. El rendimiento o la precisión de un modelo se puede

ver afectado negativamente por una metodología deficiente o por debilidades en sus datos, además de ser un requisito fundamental la validación externa del mismo, es decir, que el modelo funciona como se esperaba en nuevos pacientes similares.⁴³

Los modelos de predicción deben tener una fiabilidad y una evidencia científica avalada.³⁸ La comunidad científica debe abordar los defectos graves en el diseño, la realización y la presentación de los estudios de modelos pronósticos, y reconocer el valor clínico de una evidencia de pronóstico fiable.³⁸

En la literatura hay evidencia de que las publicaciones que describen el desarrollo y validación de modelos de predicción presentan deficiencias generalizadas en la presentación de sus informes.^{38–41,44,45} Mallet et al. revisaron 47 informes publicados en 2005 que presentaban nuevos modelos de predicción en el cáncer y encontraron que con frecuencia usaban métodos estadísticos inapropiados para el desarrollo de modelos multivariados y se caracterizaban por informes deficientes;⁴⁵ Collins et al. evaluaron 39 informes publicados antes de mayo de 2011 que describían el desarrollo de modelos predictivos de incidencia y prevalencia de diabetes tipo 2 y también describen que los informes eran generalmente deficientes, en detalles clave sobre los predictores examinados; el manejo de datos perdidos; con una estrategia de construcción de modelos mal descrita;^{40,41} Perel et al publican en 2006 una revisión sistemática de modelos pronósticos en traumatismo craneoencefálico, identificando 53 informes que incluían 102 modelos, y concluyen que gran parte de los mismos se desarrollan a partir de muestras pequeñas de pacientes, su calidad metodológica es deficiente y rara vez se validan en poblaciones externas, además de ser poco prácticos ya que no se presentan a los médicos de una manera fácil de usar.⁴⁴

Es un requisito general que se detallen los pasos de cómo se desarrolló y validó un modelo de predicción para que otros investigadores puedan reproducir esos pasos, así permitir la síntesis y la evaluación crítica de toda la información pertinente (con el fin de analizar el riesgo de sesgo y su utilidad clínica)⁴¹ y determinar las causas de un alto riesgo de sesgo o una dudosa aplicabilidad, si fuera el caso, con tal de modificarlas para mejorar dicho modelo y mejorar su

utilidad.⁴²

Por tanto, debemos realizar un análisis crítico de cómo se han construido y validado los diferentes modelos predictivos y su potencial utilización en la práctica clínica.⁴¹ Para evaluar la gran cantidad de modelos predictivos desarrollados son necesarias las revisiones sistemáticas.⁴⁶

Este análisis de los modelos predictivos en sepsis nos ayudará a determinar qué modelos de los desarrollados podemos utilizar en nuestra práctica clínica sin comprometer los resultados, basándonos en su calidad metodológica, su aplicabilidad y su validación externa.

5.2.4. Directrices internacionales.

En las últimas 2 décadas se han publicado directrices internacionales para guiar los informes de presentación de resultados de diferentes estudios de investigación. Según el diseño del estudio tenemos la guía Consolidated Standards of Reporting Trials (CONSORT) para ensayos clínicos,⁴⁷ REporting recommendations for tumour MARKer prognostic studies (REMARK) para estudios de pronóstico de marcadores tumorales,⁴⁸ The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) para estudios observacionales en epidemiología,⁴⁹ Statement for Reporting Studies of Diagnostic Accuracy (STARD) para estudios de precisión diagnóstica,⁵⁰ Genetic Risk Prediction Studies (GRIPS) para estudios de predicción de riesgo genético⁵¹ y Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) para el informe de revisiones sistemáticas y metaanálisis de ensayos clínicos aleatorizados.⁵²

Existen también diferentes herramientas de evaluación de calidad como la herramienta Cochrane Risk of Bias tool para ensayos clínicos aleatorizados,⁵³ Quality Assessment of Diagnostic Accuracy Studies (QUADAS) para estudios de validez diagnóstica⁵⁴ y la lista de verificación Quality In Prognosis Studies Tool (QUIPS) para estudios de factores pronósticos.⁵⁵

Respecto a la presentación de los informes de los estudios que desarrollan y validan modelos pronósticos, se publica en el año 2015 la declaración Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) para mejorar la transparencia en la publicación de sus resultados. La declaración TRIPOD es un conjunto de recomendaciones para la elaboración y presentación de informes de estudios que desarrollan, validan o actualizan modelos de predicción multivariante y fue desarrollada por un equipo de expertos en metodología, profesionales de la salud y editores de revistas con el objeto de mejorar la calidad de la presentación de los resultados de tales estudios. Elaboraron una lista de verificación de 22 ítems considerados esenciales para la presentación de informes transparentes de un estudio de modelo de predicción. Esta lista de verificación de 22 ítems hace referencia a los diferentes apartados de un estudio: título, resumen, introducción, métodos, resultados, discusión y otra información.⁴¹ Algunos de estos puntos son relevantes para estudios que desarrollan un modelo de predicción, otros solo son aplicables a los estudios que validan de un modelo y otros ítems se deben aplicar a ambos tipos de estudios.⁴¹

TRIPOD no es una herramienta de evaluación de la calidad de un modelo de predicción multivariante, sino que son pautas a seguir para la presentación de las investigaciones que desarrollan o validan un modelo predictivo, es decir con ellas podemos valorar la calidad de la presentación de un estudio que ha desarrollado o validado un modelo, pero no la calidad del desarrollo o validación de dicho modelo.⁴¹

En el año 2014 se publicó la declaración Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS), en la que se dan las pautas para realizar una revisión sistemática de modelos predictivos de una enfermedad en concreto. En ella se incluyen siete elementos para orientar el objetivo de la revisión, la estrategia de búsqueda y los criterios de inclusión y exclusión del estudio; y once dominios con la información a extraer para cada modelo incluido en la revisión, para conseguir una evaluación crítica del desarrollo de estos modelos. Estos once dominios son: diseño del estudio, participantes, definición del resultado o evento a predecir y su medida, predictores candidatos, tamaño de la muestra, datos perdidos, desarrollo del

modelo, rendimiento del modelo, evaluación del modelo, resultados e interpretación y discusión. La propia declaración CHARMS indica cómo debería de abordarse cada dominio, en base a la literatura científica.⁴⁶

Respecto a las recomendaciones CHARMS sobre la revisión sistemática, en primer lugar, hay que establecer si nos centramos en modelos de predicción diagnóstica o pronóstica, en función de que el modelo pronostique una situación actual o un evento futuro.⁴² En nuestro caso, nos centramos en modelos de predicción pronóstica, para estudiar modelos que pronostican mortalidad en sepsis (estimamos el riesgo de desarrollar un evento futuro –mortalidad-). El propósito de este tipo de modelos es guiar la toma de decisiones diagnósticas y terapéuticas de los médicos.

Además, debemos definir qué estudios incluimos en la revisión sistemática entre aquellos que desarrollan:^{42,46}

- un modelo sin realizar validación externa;
- un modelo y realizan validación externa en datos independientes;
- aquellos cuyo objetivo es la validación externa de un modelo, con o sin actualización de éste.

Nosotros ponemos el foco en aquellos que desarrollan un modelo predictivo, y que, podían realizar o no en la misma publicación, la validación externa del modelo. A continuación, definiremos otras cuatro preguntas sobre nuestra revisión sistemática: población objetivo a la que se aplica el modelo de predicción; resultado a predecir; periodo de tiempo en que aparece el resultado; y momento en el que utilizamos el modelo.^{42,46} Nuestro objetivo son pacientes con diagnóstico de sepsis, por lo que buscamos modelos a usar al diagnóstico o al ingreso de estos pacientes que predigan mortalidad por todas las causas desde su admisión.

Utilizamos para la revisión palabras clave recomendadas para este tipo de modelos (predictive model, prediction model, risk score, scoring system, points system and nomogram).⁴²

Respecto a los once dominios de la declaración CHARMS, que nos indican la información a extraer para poder realizar una evaluación crítica de los modelos

incluidos en nuestra revisión, vamos a detallar a continuación la información que deben contener junto con las recomendaciones para disminuir el riesgo de sesgo y aumentar la aplicabilidad clínica:^{42,46}

1. **Diseño del estudio:** es preferible un estudio de cohorte prospectivo. Las cohortes retrospectivas podrían ser adecuadas dependiendo de la calidad de los datos; los estudios transversales son solo para modelos de diagnóstico; el diseño de casos y controles no anidado es inapropiado para desarrollar un modelo predictivo; y los datos de los ensayos clínicos aleatorizados pueden tener criterios de elegibilidad restrictivos y verse afectada la generalización del modelo.
2. **Participantes:** población sobre la que se desarrolla el modelo, que debe ser representativa de la población objetivo, así como el método de selección o reclutamiento, evitando, por ejemplo, la inclusión selectiva según disponibilidad de los datos; fechas en las que se les recluta; descripción de los participantes y ámbito del estudio; debemos conocer la información de si todos los participantes se utilizaron para desarrollar o validar el modelo; y si los autores evaluaron de forma adecuada el tratamiento en el modelo predictivo.
3. **Definición del resultado o evento a predecir y su medida:** diferentes definiciones o medición de resultados pueden aumentar el riesgo de sesgo. Deben determinar el período de aparición del evento y el seguimiento de los participantes en el estudio.
4. **Predictores candidatos o variables que pueden tener una asociación con el resultado:** todos los predictores candidatos se deben indicar junto con los métodos de medición y el momento en que se midieron ya que métodos o momento de medida diferentes pueden conllevar riesgo de sesgo. Deben indicar si hubo cegamiento en la medición de predictores relacionados, en caso de que exista una posible interpretación subjetiva en la medición de éstos; y cómo se realiza el manejo de predictores continuos, debiendo evitar categorizaciones, siendo recomendables las transformaciones no lineales.
5. **Tamaño de la muestra:** número de participantes, número de eventos y número total de predictores, incluidas las transformaciones de los

predictores continuos, las variables indicadoras de los predictores categóricos y las interacciones. Se debe poder calcular la relación eventos por variable (EPV) que se recomienda sea mayor de 10.

6. **Datos perdidos:** número de participantes con datos perdidos globalmente y por cada predictor, así como el tratamiento de los datos perdidos. Se debe evitar realizar análisis de casos completos, es decir, seleccionar solamente los sujetos que tengan todos los datos de todas las variables. La imputación múltiple es el método recomendado para manejar los datos perdidos.
7. **Desarrollo del modelo:** tipo y métodos utilizados para la selección de los predictores del modelo final. Se debe evitar la selección de predictores en base al análisis univariante. Las técnicas de selección hacia adelante aumentan el riesgo de sobreajuste (overfitting) del modelo.
8. **Rendimiento del modelo:** evaluación de calibración y discriminación. El uso de medidas de clasificación y reclasificación puede producir un rendimiento demasiado optimista y sesgado.
9. **Evaluación del modelo:** método utilizado para validar el modelo. Bootstrapping es el método preferido para la validación interna del modelo.
10. **Resultados:** forma en que se presenta el modelo final (nomograma, sistema de puntos, fórmula matemática...); variables predictoras y peso de las mismas; deben informar de todos los componentes del modelo, incluida la constante en modelos de regresión logística o el riesgo basal en modelos de supervivencia; rendimiento del modelo, debemos poder calcular la probabilidad del evento para un nuevo sujeto; deben mostrar la comparación de las muestras de desarrollo y validación externa en caso de tener varios conjuntos de datos.
11. **Interpretación y discusión:** deben discutir los elementos clave en esta sección como son: resultados exploratorios, en el caso de que se necesite o se recomiende más investigación que afiance esos resultados, o confirmatorios cuando indican que el modelo es útil para la práctica; comparación con modelos anteriores y explicación de los predictores en el modelo final; generalización del modelo en otras áreas; fortalezas y limitaciones del estudio.

La declaración CHARMS, nos va a proporcionar las pautas para realizar nuestra revisión sistemática y nos indica toda la información que debemos extraer de los informes de los estudios incluidos en nuestra revisión sistemática mediante una lectura crítica de todo el artículo y todo su material suplementario.⁴²

En 2019, complementando a la declaración CHARMS,⁴⁶ surge la declaración Prediction model Risk Of Bias ASsessment Tool (PROBAST), la cual valora el riesgo de sesgo y la aplicabilidad de los modelos de predicción incluidos en una revisión sistemática.^{56,57} Ésta sería equivalente a la Cochrane Risk of Bias Tool para valorar la calidad de los ensayos clínicos contenidos en una revisión sistemática,⁵³ pero en revisiones CHARMS de modelos de predicción.⁴⁶

El sesgo generalmente se define por la presencia de un error sistemático en un estudio que conduce a resultados distorsionados o defectuosos y obstaculiza la validez interna del estudio. Muchas fuentes de sesgo identificadas en otras áreas de investigación médica también son relevantes para los estudios de modelos de predicción. Si en el desarrollo y validación de un modelo de predicción existen deficiencias en su diseño, realización o análisis, este estudio tendrá riesgo de sesgo y nos puede conducir a estimaciones incorrectas.⁵⁶

Las inquietudes con respecto a la aplicabilidad de un estudio primario pueden surgir cuando la población, los predictores o los resultados del estudio difieren de los especificados en la pregunta de revisión. Tales preocupaciones pueden surgir cuando, por ejemplo, los participantes en el estudio del modelo de predicción provienen de un entorno médico diferente de la población definida en la pregunta de revisión; un ejemplo sería un estudio que inscribe a pacientes de un entorno hospitalario mientras que la pregunta de revisión se relaciona específicamente con pacientes en atención primaria. Cuando los criterios de elegibilidad, los predictores y los resultados de los estudios primarios coinciden y se ajustan con la pregunta de investigación de la revisión sistemática, no surgirán preocupaciones con respecto a la aplicabilidad.⁵⁶

PROBAST valora el riesgo de sesgo a través de cuatro dominios (participantes, predictores, resultado y análisis) y la aplicabilidad mediante tres de los mismos (participantes, predictores y resultado). El dominio "Participantes" se refiere a la selección de la muestra del estudio, "Predictores" evalúa los predictores

candidatos que forman parte del modelo, “Resultado” analiza la variable que estamos prediciendo y “Análisis” aborda los métodos estadísticos. Cada dominio se califica como “bajo”, “alto” o “incierto”, ya sea para riesgo de sesgo o para la preocupación respecto a la aplicabilidad. Para poder emitir este juicio de alto, bajo o incierto riesgo de sesgo o de preocupación respecto a la aplicabilidad, de una forma objetiva, la misma guía PROBAST indica unas directrices y un total de 20 preguntas que deben ser contestadas para valorar los diferentes dominios.^{56,57}

Finalmente, se emite una valoración global, tanto para el riesgo de sesgo como para la aplicabilidad, utilizando el principio “*the worst score counts*” o “la peor puntuación cuenta”, significando que un ítem negativo en un dominio hace que la valoración global de ese apartado, ya sea riesgo de sesgo o aplicabilidad, sea negativa, y por lo tanto su uso se desaconseje. La calificación de cada dominio se realiza objetivamente bajo las recomendaciones PROBAST, las cuales, al igual que CHARMS,⁴⁶ se basan en la literatura científica.^{56,57}

Existen revisiones sistemáticas de modelos predictivos realizadas con anterioridad a la publicación de la declaración CHARMS, y estos estudios llevaron a cabo, generalmente, un análisis descriptivo general de los resultados de cada artículo incluido en la revisión, en lugar de un análisis extenso de cada uno de los modelos predictivos.^{40,42,44,45} Para evaluar un modelo de predicción, es clave conocer todos los detalles de su desarrollo, validación y forma de utilización, y lo conseguiremos realizando un análisis crítico extenso de cada modelo. Para realizar ésta evaluación crítica detallada existe, además de las guías CHARMS⁴⁶ y PROBAST^{56,57}, un documento explicativo para realizar una revisión CHARMS de modelos de predicción.⁴²

A pesar de que estas guías que compendian las recomendaciones para estudios que desarrollan un modelo de predicción son relativamente recientes, los modelos desarrollados con anterioridad a la publicación de las mismas se deben adecuar a las recomendaciones internacionales, ampliamente discutidas en la literatura científica previa.³⁹ Estas guías pueden ayudar a identificar los aspectos en los que existen deficiencias en estos estudios que han desarrollado un modelo de predicción y mejorarlos con el objeto de conseguir un bajo riesgo de sesgo y

una adecuada aplicabilidad, recomendando de esta forma su incorporación a la práctica clínica habitual.





6.JUSTIFICACIÓN





La sepsis tiene una elevada tasa de mortalidad, de hasta un 25%,⁴ y es fundamental determinar precozmente la gravedad de la misma para la instauración precoz de las medidas diagnóstico-terapéuticas ajustadas a este riesgo pronóstico. Para categorizar la gravedad de la sepsis debemos utilizar modelos pronósticos cuyo desarrollo y validación se adapten a las recomendaciones internacionales. Sin embargo, y hasta donde nosotros conocemos, no se ha realizado una revisión sistemática de modelos pronósticos para esta enfermedad con la aplicación de PROBAST. En consecuencia, el objetivo de este estudio fue la realización de una revisión sistemática de estas características, proporcionando un documento crítico que determine la idoneidad de la integración de los modelos pronósticos existentes para pacientes sépticos en la práctica clínica.





7.OBJETIVOS





El objetivo principal de este estudio fue:

- Realizar una revisión sistemática de estudios que desarrollen y validen modelos de predicción de mortalidad en pacientes con sepsis.

Los objetivos secundarios fueron los siguientes:

- Conocer los modelos predictivos de mortalidad en sepsis que se han desarrollado hasta el 17 de septiembre de 2019.
- Revisar la metodología del desarrollo de estos modelos predictivos.
- Evaluar el riesgo de sesgo en el desarrollo de los mismos.
- Ver qué modelos de los analizados están validados externamente.
- Determinar la aplicabilidad clínica de estos modelos para estimar el riesgo de mortalidad sobre pacientes de forma individual.





8.MATERIALES Y MÉTODOS





8.1. Protocolo y registro.

Esta revisión ha sido registrada en la base de datos PROSPERO (CDR42018111121), en donde se especificaron a priori los métodos en los que ésta se basaría. Además, esta revisión se basa en las declaraciones CHARMS y PROBAST.^{46,56,57}

8.2. Criterios de inclusión y exclusión.

Realizamos una revisión sistemática de artículos científicos que, sobre un conjunto de datos propios en pacientes con diagnóstico de sepsis, desarrollaron un modelo predictivo, de al menos dos predictores, de mortalidad, lo cual corresponde a un modelo pronóstico. Éstos además podían realizar en la publicación la validación interna o externa del modelo. La validación interna es la que se hace en el conjunto de datos utilizado para desarrollar el modelo y la validación externa es la que se realiza en cualquier otro conjunto de datos. Se consideraron los modelos que incluían muestras de pacientes con sepsis en general, es decir, sin ser causadas por un patógeno específico. La definición de sepsis utilizada en nuestro estudio era la vigente en el momento de desarrollar cada uno de los modelos. Nuestro objetivo es obtener modelos que puedan utilizarse en el momento del diagnóstico o del ingreso del paciente por enfermedad séptica, para predecir mortalidad por todas las causas, independientemente del momento en que ocurra el fallecimiento desde la situación basal. Los diseños que se corresponden con todas estas características son los estudios de cohortes, ya sean prospectivos o retrospectivos, pues vamos a predecir un evento futuro, que en nuestro caso es la mortalidad.⁴⁶

Se excluyeron artículos de revisión o que describen el desarrollo de un modelo sin usar datos propios (por ejemplo datos simulados), resúmenes de congresos, estudios que analizan el valor de nuevos predictores en modelos existentes, artículos que realizan únicamente una validación externa de un modelo predictivo ya publicado o que no desarrollen un modelo de riesgo multivariante, trabajos que se desarrollaron sobre población pediátrica o quirúrgica, artículos

que se refieran a bacteriemia u otra entidad infecciosa no definitoria de sepsis, o los que se refieran a un microorganismo concreto causante de la infección. También se excluyeron los trabajos que aplican técnicas de Machine Learning, excepto aquellas técnicas analíticas que utilizaron modelos de regresión logística. Los otros tipos de modelos no son adecuados para la declaración CHARMS y la lista de verificación PROBAST.

8.3. Fuente de datos

Se analizaron las bases de datos MEDLINE (a través de Pubmed), Scopus y EMBASE, desde su creación hasta el 17 de septiembre de 2019. La creación de estas bases de datos data de 1947 para EMBASE, y de 1966 para MEDLINE y Scopus.

8.4. Búsqueda y selección de estudios

Las bases de datos fueron analizadas utilizando palabras clave referidas a varios términos: modelos de predicción (predictive model, prediction model, risk score, scoring system, points system and nomogram), tipo de pacientes (sepsis) y outcome (mortality, death and prognosis). Estas palabras clave fueron buscadas en el título, en el resumen o en tesoro de cada base de datos analizada (MESH y EMTREE). Los artículos fueron restringidos a aquellos publicados en inglés o en castellano con resumen (idiomas que podíamos comprender los autores del trabajo). Se analizaron las referencias de todos los artículos que podían ser relevantes a nuestra pregunta de investigación. Las ecuaciones de búsqueda se dan en el Apéndice S1 (Anexo).

Dos revisores cegados, independientemente, analizaron los títulos y resúmenes para determinar qué artículos científicos podían ser potencialmente incluidos. Los resultados fueron puestos en común y los desacuerdos fueron resueltos por consenso. El mismo procedimiento fue aplicado al texto completo de los artículos asociados con esos títulos y resúmenes, al igual que para las referencias de los

artículos incluidos. Estos revisores eran expertos en modelos de predicción y/o enfermedad séptica. No se realizaron búsquedas manuales para analizar la documentación gris, al ser una temática compleja, propia de las revistas científicas.

8.5. Extracción de datos

De los artículos finalmente incluidos en la revisión, los mismos dos revisores cegados, independientemente extrajeron la información requerida por la declaración CHARMS.⁴⁶ Ésta se agrupa en 11 dominios: 1) Diseño del estudio; 2) Participantes: descripción de los participantes, fechas y tratamientos; 3) Resultado o evento a predecir: definición del resultado a predecir y período de aparición del mismo; 4) Predictores candidatos: cantidad y tipo, forma y momento de medición y manejo de predictores continuos; 5) Tamaño de la muestra: número de participantes, número de eventos y la relación eventos por variable; 6) Datos perdidos: número de participantes con datos perdidos globalmente y por cada predictor, y tratamiento de los datos perdidos; 7) Desarrollo del modelo: tipo y métodos utilizados para la selección de predictores; 8) Rendimiento del modelo: calibración y discriminación; 9) Evaluación del modelo: método utilizado; 10) Resultados: forma de presentación del modelo final (nomograma, sistema de puntos, fórmula matemática...), peso de las diferentes variables y comparación de muestras de desarrollo y validación externa cuando proceda; y 11) Interpretación y discusión: comparación con estudios similares, generalización del modelo, fortalezas y limitaciones.⁴⁶Toda esta información será presentada de forma tabular para mejorar la visualización de todas las características de los modelos.^{58,59}

8.6. Análisis de la calidad.

Después de haber extraído la información indicada en CHARMS,⁴⁶ de forma duplicada también, se aplicó la herramienta PROBAST, la cual valora tanto el riesgo de sesgo como la aplicabilidad de los modelos predictivos.^{56,57} El riesgo de sesgo se divide en cuatro dominios: participantes, predictores, resultado y análisis; y la aplicabilidad en tres dominios: participantes, predictores y resultado. El dominio *Participantes* se refiere a la selección de la muestra del estudio, *Predictores* evalúa los predictores candidatos que forman parte del modelo, *Resultado* analiza la variable que estamos prediciendo y *Análisis* aborda los métodos estadísticos. La propia declaración PROBAST indica de forma objetiva cómo valorar cada uno de dichos dominios. Los siete dominios se pueden categorizar en “alto”, “bajo” o “incierto”, ya sea riesgo de sesgo o preocupación respecto a la aplicabilidad. Tras valorar los dominios de forma independiente, se emite una valoración global para los dos bloques, utilizando el principio “la peor puntuación cuenta”, significando que un ítem negativo en un dominio hace que la valoración global sea negativa, y por lo tanto el uso de ese modelo se desaconseje. Por ejemplo, si en el riesgo de sesgo hemos obtenido un nivel de bajo riesgo para todos los dominios excepto para el de Análisis en el que se obtiene “alto” el riesgo de sesgo, la valoración global del riesgo de sesgo será de “alto riesgo de sesgo” (domina la peor puntuación). El riesgo de sesgo determina si se han realizado errores sistemáticos en el estudio, mientras que la aplicabilidad indica si el modelo se ajusta a la pregunta de investigación de la revisión sistemática,^{56,57} que en este caso es la determinación del riesgo de muerte en un paciente con sepsis.

8.7. Análisis estadístico

Se ha realizado un análisis descriptivo tanto de los dominios de CHARMS como de PROBAST,^{46,56,57} pues no tiene sentido realizar un meta-análisis en este tipo de estudio. Además, para comparar nuestros resultados con otros estudios similares, buscamos revisiones sistemáticas que utilizan la lista de verificación PROBAST en otras enfermedades, mediante el análisis de sus citas (Scopus),

obteniendo el porcentaje de bajo riesgo de sesgo y baja preocupación respecto a su aplicabilidad para todos los dominios. También empleamos el portal Scopus para analizar los artículos que habían referenciado a los trabajos incluidos en esta revisión con el objeto de ver si habían sido validados externamente.





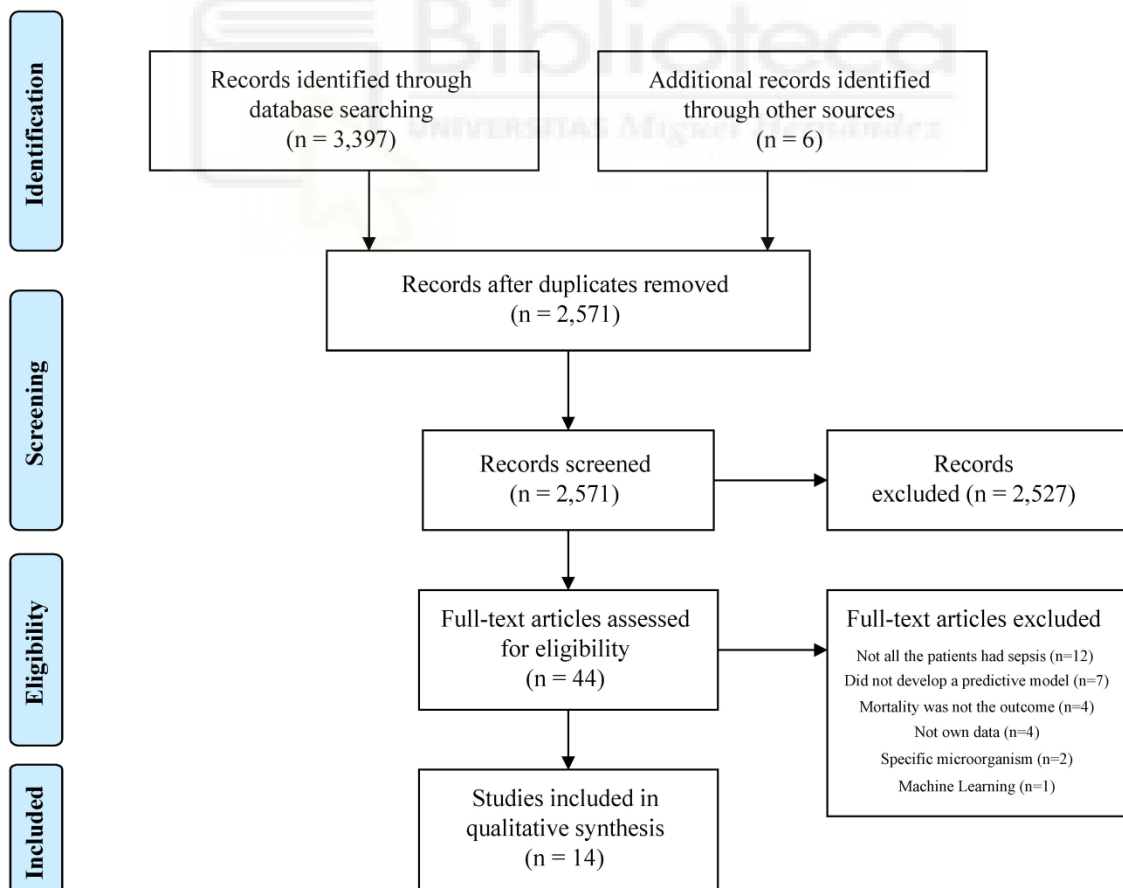
9.RESULTADOS





En la Figura 1 se observa el esquema de la revisión sistemática. Un total de 3.397 artículos fueron cribados por título y resumen en las bases de datos analizadas (485 en MEDLINE, 1.254 en Scopus y 1.658 en EMBASE) y 44 artículos fueron seleccionados para analizar a texto completo.^{16,60-102} De ellos, 12 fueron excluidos porque no todos los pacientes tenían diagnóstico de sepsis,⁶⁰⁻⁷¹ 7 no desarrollaron un modelo predictivo,⁷²⁻⁷⁸ en 4 el evento a predecir no era la mortalidad,^{16,79-81} en otros 4 no se disponía de datos propios,⁸²⁻⁸⁵ 2 valoraban únicamente un microorganismo específico y no sepsis en general,^{86,87} y 1 aplicaba técnicas de Machine Learning,⁸⁸ no contempladas en esta revisión. Finalmente, 14 cumplieron todos los criterios de inclusión y ninguno de exclusión,⁸⁹⁻¹⁰² a los que se les extrajo toda la información de acuerdo a la declaración CHARMS y se valoró el riesgo de sesgo y la aplicabilidad según la guía PROBAST.

Figura 1:



En las Tablas S1-5 (Anexo) se muestra de forma completamente detallada lo referido a los 11 dominios de CHARMS, y en la tabla 10 un resumen de estos resultados obtenidos. Tras un análisis descriptivo observamos que la mitad de los estudios tenían un diseño prospectivo y la otra mitad retrospectivo (*Diseño del estudio*). En lo referente a *Participantes*, destacamos que generalmente éstos fueron pacientes adultos con diagnóstico de sepsis o sepsis severa, siendo a menudo el origen urgencias o la unidad de cuidados intensivos. Además, nos gustaría señalar que hubo 9 estudios que realizaron validación externa sobre otro conjunto de datos.^{92-96,98,99,101,102} Respecto al *Tratamiento*, 5 autores lo valoraron como predictor candidato.^{95,97,98,100,101} Las fechas de recogida de datos oscilaron entre el año 2000 y 2016. El *Evento a predecir* fue siempre mortalidad por todas las causas, indicando la aparición de la misma en periodo de hospitalización o de ingreso en la unidad de cuidados intensivos. En algunos estudios además se indicó el tiempo para la predicción, con un rango de 28-30 días.^{89-91,93,94,96}

En el dominio *Predictores candidatos* hubo gran variabilidad, pudiéndose clasificar por norma general en demográficos, comorbilidades, scores pronósticos (modelos predictivos de gravedad en las Unidades de Cuidados Intensivos), constantes vitales, datos de laboratorio y características y tratamiento de la infección. La mayoría de estos predictores eran obtenidos en el momento del ingreso, pero hubo sólo 7 artículos que tenían todas las variables necesarias para aplicar el modelo en el momento del diagnóstico (situación basal) es decir, se disponía de toda la información de los predictores en la situación basal.^{89,91-94,99,102} Respecto al manejo de predictores continuos, la mayoría utilizó categorizaciones.^{89,93-95,97-100,102} En el dominio *Tamaño muestral* se observó un rango de eventos por variable que osciló entre 0,57 y 572,28, aunque hubo 3 trabajos en los que no pudimos calcularlo, ya sea por desconocer el número de eventos o de predictores.^{95,98,102} En los trabajos que realizaron validación del modelo construido sobre otro conjunto de datos,^{92-96,98,99,101,102} el número de eventos osciló entre 9 y 56,579, aunque hubo 2 trabajos que no aportaron esta información en la publicación.^{98,102}

En el dominio *Datos perdidos* se observó que la mayoría no aportaban información al respecto,^{89-92,95-97,100,101} tres utilizaron el análisis de casos

completos,^{93,94,99} y sólo 1 utilizó imputaciones múltiples, pero únicamente con un predictor.⁹⁸ Respecto al *Desarrollo del modelo*, se aprecia que todos los trabajos estimaron modelos de regresión logística binaria,⁸⁹⁻¹⁰² habiendo 9 autores que se basaron en el análisis univariante para seleccionar los predictores para su inclusión en el modelo multivariante.^{89-91,93,96,98-100,102}

En *Rendimiento del modelo* destacamos que en todos los modelos se calculó la discriminación a través del área bajo la curva ROC (receiver operating characteristic).⁸⁹⁻¹⁰² La calibración fue realizada principalmente por el test de Hosmer-Lemeshow y/o gráficamente, aunque hubo 3 modelos en los que no se abordó este aspecto.^{90,91,94} En *Evaluación del modelo* sólo en 3 modelos se aplicaron técnicas de remuestreo (bootstrapping),^{89,98,102} y en otros 2, aunque no se aplicaron este tipo de técnicas, sí se tuvo en cuenta el sobreajuste (overfitting).^{92,93} Los valores obtenidos de discriminación y calibración para cada modelo, quedan reflejados en la Tabla 9, donde observamos que el área bajo la curva ROC osciló entre 0,59 y 0,955 y que no se encontró ninguna falta de calibración.

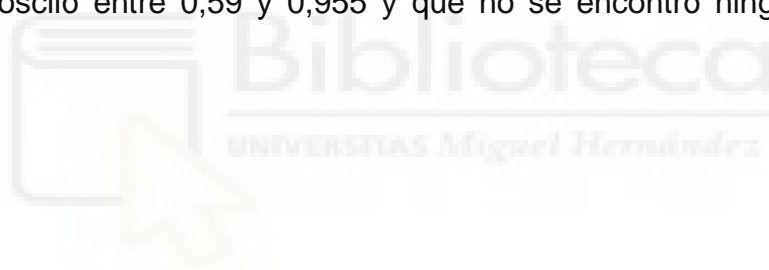


Tabla 9: Discriminación y calibración de los modelos predictivos incluidos en nuestra revisión sistemática.

Referencias	Discriminación (AUC con IC 95%)	Calibración
García-Lamberechts et al., 2018 ⁷⁷	0,79 (0,72-0,85)	p=0,496 (test Hosmer-Lemeshow)
Jiang et al., 2018 ⁷⁸	0,955 (0,901-1)	No realizada
Samsudin et al., 2018 ⁷⁹	0,79 (0,72-0,86)	No realizada
Schwarzkopf et al., 2018 ⁸⁰	0,737 (0,734-0,74) (derivación) 0,736 (0,733-0,739) (validación) 0,739 (0,737-0,743) (validación)	Tasa de mortalidad en decil de riesgo más bajo y más alto (derivación): 0,11-0,78 Tasa de mortalidad en decil de riesgo más bajo y más alto (validación): 0,1-0,76 Tasa de mortalidad en decil de riesgo más bajo y más alto (validación): 0,1-0,77 Gráficos de calibración (los riesgos observados y esperados fueron bastante similares)
Bewersdorf et al., 2017 ⁸¹	0,83 (0,79-0,88) (derivación) 0,80 (0,73-0,90) (validación)	p=0,42 (test Hosmer-Lemeshow para la muestra de derivación)
Fang et al., 2017 ⁸²	0,853 (0,769-0,915) (derivación) 0,789 (0,641-0,896) (validación)	No realizada
Ford et al., 2016 ⁸³	0,801 (derivación) 0,765 (validación)	p=0,37 (test Hosmer-Lemeshow para la muestra de derivación) p=0,56 (test Hosmer-Lemeshow para la muestra de validación) Gráficos de calibración (los riesgos observados y esperados fueron bastante similares)
Seo et al., 2016 ⁸⁴	0,8173 (0,7605–0,8741) (derivación) 0,7537 (0,6563–0,8512) (validación)	p=0,61 (test Hosmer-Lemeshow para la muestra de derivación) Gráficos de calibración (los riesgos observados y esperados fueron bastante similares)

Referencias	Discriminación (AUC con IC 95%)	Calibración
Mohan et al., 2015 ⁸⁵	0,8422	No proporcionado
Osborn et al., 2014 ⁸⁶	0,736 (derivación) 0,748 (validación)	p=0,373 (test Hosmer-Lemeshow para la muestra de derivación) p=0,937 (test Hosmer-Lemeshow para la muestra de validación) Gráficos de calibración (los riesgos observados y esperados fueron bastante similares)
Sivayoham et al., 2013 ⁸⁷	0,712 (0,64-0,78) (derivación) 0,68 (0,6-0,77) (validación)	p=0,6 (test Hosmer-Lemeshow para la muestra de derivación) p=0,3 (test Hosmer-Lemeshow para la muestra de validación)
Granja et al., 2013 ⁸⁸	0,66 (0,62-0,69) (Modelo 1) 0,59 (0,55-0,63) (Modelo 2) 0,72 (0,68-0,75) (Modelo 3) 0,81 (0,78-0,84) (Modelo 4) 0,84 (0,81-0,87) (Modelo 5)	p=0,297 (test Hosmer-Lemeshow para el Modelo 1) p=0,630 (test Hosmer-Lemeshow para el Modelo 2) p=0,678 (test Hosmer-Lemeshow para el Modelo 3) p=0,588 (test Hosmer-Lemeshow para el Modelo 4) p=0,368 (test Hosmer-Lemeshow para el Modelo 5)
Lagu et al., 2011 ⁸⁹	0,78 (derivación) 0,693 (0,649-0,775) (validación)	Gráficos de calibración para la muestra de derivación (los riesgos observados y esperados fueron bastante similares) p=0,11 (test Hosmer-Lemeshow para la muestra de validación)
Tabak et al., 2007 ⁹⁰	No proporcionado	Gráficos de calibración (los riesgos observados y esperados fueron bastante similares)

Abreviaturas: AUC, área bajo la curva ROC (receiver operating characteristic); IC, intervalo de confianza. Cuando el IC no es proporcionado en el artículo original, no podemos escribirlo en la tabla.

En *Resultados* llama la atención que sólo dos grupos de autores tenían todos los resultados que se habían propuesto en la metodología,^{90,98} siendo la falta de la constante del modelo de regresión la omisión más común.^{89,91,93,94,96,97,99-102} Otro punto importante es la forma de presentar el modelo, siendo la opción mayoritaria el sistema de puntos (n=7),^{89,93-95,97-99} seguido de la fórmula matemática (coeficientes de regresión, n=6),^{90-92,100-102} y de nomograma (n=1).⁹⁶ Por último, en *Interpretación y discusión*, destacamos que la mayoría (13/14) indicaron en su discusión que había que tener cautela en la utilización del modelo (resultados exploratorios), pero uno de ellos recomendaba de forma directa la implementación del modelo en la práctica clínica habitual.¹⁰² Un total de 12 trabajos sí discutió tanto los predictores encontrados en los modelos, como la comparación de su modelo con otros previos ya publicados.^{89-93,96-102} Las fortalezas y limitaciones sí fueron valoradas por todos los trabajos.⁸⁹⁻¹⁰² Finalmente, hubo dos trabajos que no discutieron la generalización del modelo a otras áreas geográficas.^{98,102}



Tabla 10: Resumen resultados del análisis CHARMS.

Resultados CHARMS		
Diseño del estudio	50% prospectivos y 50% retrospectivos	
Participantes	<ul style="list-style-type: none"> - Adultos con sepsis. - Origen: urgencias o UCI. - Datos recogidos de 2000 a 2016. 	<ul style="list-style-type: none"> - 9 estudios realizaron validación externa. - 5 estudios valoran el tratamiento como predictor.
Evento	Mortalidad por todas las causas.	En 6 estudios establecen un rango de 28-30 días.
Predictores candidatos	<ul style="list-style-type: none"> - Demográficos, comorbilidades, clínicos, tratamiento. - 7 estudios tenían todas las variables al diagnóstico. 	<ul style="list-style-type: none"> - 9 estudios realizan categorizaciones para los predictores continuos.
Tamaño muestral	Rango de EPV de 0.57 a 572.28.	Los que hacen validación tienen de 9 a 56 eventos.
Datos perdidos	<ul style="list-style-type: none"> - 3 autores utilizan análisis de casos completos. - 1 autor realiza imputaciones múltiples. 	<ul style="list-style-type: none"> - 9 estudios no aportan información.
Desarrollo del modelo	Todos los trabajos modelos de regresión logística binaria.	<ul style="list-style-type: none"> - 9 autores utilizan el análisis univariante para la selección de variables.
Rendimiento del modelo	Todos calcularon discriminación a través del área bajo la curva ROC.	<ul style="list-style-type: none"> - Calibración realizada con el test Hosmer-Lemeshow y/o gráficamente. - En 3 modelos no se abordó este aspecto.
Evaluación del modelo	3 modelos técnicas de remuestreo.	2 tuvieron en cuenta el sobreajuste.
Resultados	<ul style="list-style-type: none"> - 2 autores obtuvieron todos los resultados propuesto s. - Falta más común la constante del modelo de regresión. 	<ul style="list-style-type: none"> - 7 sistemas de puntos - 6 fórmulas matemáticas - 1 nomograma.
Interpretación y discusión	<ul style="list-style-type: none"> - 13 recomiendan cautela para utilizar el modelo. - 12 realizaron comparación con otros modelos. 	<ul style="list-style-type: none"> - Todos valoran fortalezas y limitaciones. - 2 trabajos no discuten generalización en otras áreas geográficas.

CHARMS: CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies; UCI: unidad de cuidados intensivos; EPV: eventos por variable; ROC: receiver operating characteristic.

El análisis del PROBAST queda reflejado en la Tabla 11 y de forma más detallada en el Apéndice S2 (Anexo). En dicha tabla observamos los dominios tanto para el *riesgo de sesgo* como para la *aplicabilidad*. En lo referente a los dominios comunes entre estos dos aspectos (participantes, predictores y resultado), se obtuvo la misma valoración tanto para el riesgo de sesgo como para la aplicabilidad. En *Participantes* se consensuó una puntuación positiva para 10 estudios (71,4%),^{89,92-96,98-100,102} y en uno faltaba información para determinar el ítem.¹⁰¹ Para *Predictores* la mitad de los trabajos alcanzaron estándares de calidad satisfactorios,^{89,91-93,96,99,102} y en 3 no estuvo claro.^{90,94,101} En *Resultado* la puntuación máxima se encontró en 12 estudios (85,7%),^{89,91-93,95-102} no quedando claro en el resto.^{90,94} Valorando globalmente estos tres ítems, esto condujo a un total de 6 estudios con baja preocupación respecto a su aplicabilidad.^{89,92,93,96,99,102}

El dominio del *Análisis*, únicamente presente en la valoración del riesgo de sesgo, resultó que en 13 de los 14 estudios la valoración era negativa,^{89-91,93-102} y en el restante no se pudo valorar.⁹² Hemos de tener en cuenta que este dominio tiene un total de 9 cuestiones y con que una de ellas sea valorada negativamente, el dominio es cuantificado como de alto riesgo de sesgo (“la peor puntuación puntúa”). Entre los 14 trabajos incluidos, hubo un único trabajo con 8 cuestiones valoradas positivamente de las 9 que se valoran (falló en indicar cómo manejó los datos perdidos, no quedando claro este ítem),⁹² otro tuvo 7 cuestiones cuantificadas satisfactoriamente y las otras dos negativas (en el manejo de los datos perdidos por juntar datos perdidos con la categoría de referencia; y en la utilización del análisis univariante para seleccionar los predictores del modelo),¹⁰² y tres trabajos tuvieron seis cuestiones positivas y tres con valoración negativa o no aclarada: 1) Ford et al., 2016: utilizó categorizaciones en los predictores continuos y no tuvo en cuenta el sobreajuste en la validación interna, y en datos perdidos no había forma de valorarlo.⁹⁵; 2) Osborn et al., 2014: no realizó el manejo de predictores continuos adecuadamente y empleó el análisis univariante para seleccionar los predictores del modelo, y no pudimos valorar la cuestión de los datos perdidos.⁹⁸ y 3) Lagu et al., 2011: no tuvo en cuenta el sobreajuste y hubo dos ítems de los que no teníamos información para valorarlos, siendo el primero de ellos los datos

perdidos y el segundo la dudosa utilización de todos los pacientes para desarrollar el modelo, pues al hacer exclusiones después de iniciar el seguimiento, no queda claro.¹⁰¹



Tabla 11: Riesgo de sesgo y preocupación respecto a la aplicabilidad de los estudios incluidos que evalúan modelos pronósticos multivariados que predicen mortalidad en pacientes con sepsis. (PROBAST).

Estudio	ROB				Aplicabilidad			Valoración global	
	Participantes	Predictores	Resultado	Análisis	Participantes	Predictores	Resultado	ROB	Aplicabilidad
García-Lamberchts et al., 2018 ⁷⁷	+	+	+	-	+	+	+	-	+
Jiang et al., 2018 ⁷⁸	-	?	?	-	-	?	?	-	-
Samsudin et al., 2018 ⁷⁹	-	+	+	-	-	+	+	-	-
Schwarzkopf et al., 2018 ⁸⁰	+	+	+	?	+	+	+	?	+
Bewersdorff et al., 2017 ⁸¹	+	+	+	-	+	+	+	-	+
Fang et al., 2017 ⁸²	+	?	?	-	+	?	?	-	?
Ford et al., 2016 ⁸³	+	-	+	-	+	-	+	-	-
Seo et al., 2016 ⁸⁴	+	+	+	-	+	+	+	-	+
Mohan et al., 2015 ⁸⁵	-	-	+	-	-	-	+	-	-
Osborn et al., 2014 ⁸⁶	+	-	+	-	+	-	+	-	-
Sivayoham et al., 2014 ⁸⁷	+	+	+	-	+	+	+	-	+
Granja et al., 2013 ⁸⁸	+	-	+	-	+	-	+	-	-
Lagu et al., 2011 ⁸⁹	?	?	+	-	?	?	+	-	?
Tabak et al., 2007 ⁹⁰	+	+	+	-	+	+	+	-	+

Abreviaturas: PROBAST, Prediction model Risk Of Bias ASsesment Tool; ROB, riesgo de sesgo.

+: bajo ROB/baja preocupación respecto a la aplicabilidad; -: alto ROB/alta preocupación respecto a la aplicabilidad; ?: incierto ROB/incierta preocupación respecto a la aplicabilidad.

En los estudios de otras enfermedades consultados,^{58,103–107} cabe destacar que el análisis fue el dominio que presentó más riesgo de sesgo, ya que en varias revisiones ninguno o muy pocos trabajos de los incluidos lo realizó correctamente. En cambio, el resto de dominios, tanto en riesgo de sesgo como aplicabilidad, fueron bastante satisfactorios, superando el 70% casi todas las revisiones (Tabla 12).

Tabla 12: Bajo riesgo de sesgo y baja preocupación respecto a la aplicabilidad (%) de otras revisiones sistemáticas que evalúan modelos predictivos multivariantes para otras enfermedades. (PROBAST).

Estudios	Enfermedad	ROB				Aplicabilidad			Valoración global	
		Participantes	Predictores	Resultado	Análisis	Participantes	Predictores	Resultado	ROB	Aplicabilidad
Nuestra revisión	Septic	71,4	50	85,7	0	71,4	50	85,7	0	42,9
Palazón-Bru et al., 2020 ⁶⁶	Cáncer orofaríngeo	83,3	100	100	0	83,3	100	100	0	83,3
Du et al., 2020 ⁹¹	Salud bucodental	33,3	79,2	70,8	8,3	N/A	N/A	N/A	4,2	N/A
Di Tanna et al., 2020 ⁹²	Insuficiencia cardíaca	75,0	62,5	65,0	20,0	80,0	97,5	82,5	17,5	87,5
Bellou et al., 2019 ⁹³	EPOC	82,8	96,6	98,5	2,5	N/A	N/A	N/A	1,7	N/A
Carrillo-Larco et al., 2019 ⁹⁴	Diabetes mellitus	83,3	50	50	0	83,3	83,3	83,3	0	83,3
Carrillo-Larco et al., 2019 ⁹⁵	Enfermedades cardiovasculares	100	100	87,5	0	50	100	100	0	50

Abreviaturas: EPOC, Enfermedad Pulmonar Obstructiva Crónica; N/A, no realizado; PROBAST, Prediction model Risk Of Bias ASsesment Tool; ROB, riesgo de sesgo.

De los trabajos encontrados en nuestra revisión sistemática, sólo 3 modelos predictivos fueron sometidos a estudios de validación externa: los liderados por Osborn,⁹⁸ Sivayoham,⁹⁹ y Granja.¹⁰⁰ El primero tuvo cuatro estudios de validación externa,¹⁰⁸⁻¹¹¹ y los otros uno.^{112,113}



10. DISCUSIÓN





10.1. Fortalezas y limitaciones.

La fortaleza principal de este trabajo es el hecho de que utilizar un diseño sistemático, en el que dos personas de forma independiente revisan toda la información (alrededor de 2.500 artículos científicos) y evaluar los datos mediante herramientas desarrolladas por expertos en modelos de predicción,^{46,56,57} disminuye el riesgo de errores sistemáticos.

Podría ser que, a pesar de haber realizado una búsqueda exhaustiva y sistemática en las bases de datos consultadas, no hubiéramos detectado algún trabajo que sí desarrollara un modelo de predicción con los criterios de selección prefijados. También podría ser que en otras bases de datos sí existieran otras publicaciones con estos criterios no indexadas en MEDLINE, EMBASE o Scopus. Además, podría haber publicaciones sí indexadas en dichas bases de datos, pero en otros idiomas que los autores no podemos comprender. No obstante, la gran mayoría de la literatura científica está publicada en inglés y, de hecho, todos nuestros artículos están en dicho idioma, salvo uno que se encuentra tanto en castellano como en inglés.⁸⁹ Finalmente, este tipo de revisión es una crítica metodológica de los modelos predictivos, la cual no incluye los modelos en sí, aunque sí muchos detalles de los mismos. En consecuencia, si el lector quiere recibir más información al respecto, debería acudir a la referencia original del modelo, donde los autores deberían haber indicado el modelo y cómo aplicarlo sobre un paciente real.

10.2. Comparación con la literatura existente.

A la hora de comparar nuestros resultados (riesgo de sesgo y aplicabilidad), debemos de analizar revisiones sistemáticas con metodologías similares que hayan utilizado el PROBAST, encontrando gran variabilidad de tipos de pacientes y en el evento a medir.^{58,103–107} Se han indicado los porcentajes de bajo riesgo y los de baja preocupación respecto a la aplicabilidad de todos estos estudios en la Tabla 12. En ella podemos observar que los resultados obtenidos en la sepsis para todos los dominios, salvo el de *Predictores*, se encuentran en

el rango de los otros trabajos. Esto podría ser debido a que los modelos de sepsis (Tablas S1-5, Anexo) tienen parámetros de laboratorio e información sobre el tratamiento que podría no estar disponible en el momento de hacer la predicción. No obstante, es algo difícil de comparar, pues son enfermedades y predictores muy diferentes.

10.3. Implicaciones en la investigación y en la práctica clínica.

Dados los resultados encontrados en nuestra revisión, pensamos que hay dos puntos de mejora de cara a la investigación, ya que se ha encontrado alto riesgo de sesgo y alta preocupación respecto a la aplicabilidad en un porcentaje importante de los modelos estudiados. En primer lugar, si se desarrolla un nuevo modelo predictivo, éste deberá basarse tanto en la declaración CHARMS como en la de PROBAST,^{46,56,57} de forma que el modelo pueda implementarse en la práctica clínica habitual, aunque es importante señalar que la decisión de implementar un modelo de predicción de riesgo en la práctica clínica no se basa únicamente en el informe de elementos de una lista de verificación o en el riesgo de sesgo. Hay otros aspectos importantes, como su precisión (discriminación y calibración) y facilidad de uso, además de que haya sido validado externamente varias veces. En segundo lugar, se puede mejorar la utilización de las bases de datos de los estudios analizados (Tablas S1-5, Anexo), seleccionando correctamente los pacientes y los predictores, teniendo en cuenta los aspectos de la definición del evento a predecir y aplicando las técnicas estadísticas recomendadas. De esta forma, sin el esfuerzo importante de recoger datos de un gran número de pacientes de nuevo, se pueden corregir las deficiencias encontradas.

Una idea interesante sería comparar los parámetros de discriminación y calibración entre los estudios con alto y bajo riesgo de sesgo, determinando si esto es un factor que pueda alterar la precisión del modelo. Nosotros, al no haber obtenido ningún trabajo con bajo riesgo de sesgo, no hemos podido realizar esta comparativa, pero sería una buena propuesta para otra revisión similar a la nuestra.

Otra futura línea de investigación con utilidad para la práctica clínica, sería realizar una revisión de modelos predictivos, valorando únicamente validaciones externas realizadas sobre pacientes con sepsis. Esto es una revisión distinta a la planteada, que se corresponde con el subtipo 3 de modelos valorados en la declaración CHARMS: “Estudios de validación externa de modelos de riesgo con o sin actualización del modelo”.⁴⁶ Los trabajos en esta nueva revisión podrían incluir validaciones de modelos desarrollados en pacientes con sepsis, como los incluidos en nuestro trabajo, o modelos realizados en otro tipo de poblaciones, pero validados en pacientes con sepsis. Esta nueva revisión complementaría la desarrollada en este trabajo.

En lo referente a la práctica clínica, debemos de ser muy cautos, pues únicamente un modelo podría cumplir los estándares de calidad del PROBAST.⁹² No obstante, sus propios autores indican que el modelo desarrollado no debería de aplicarse a la práctica clínica, ya que utiliza datos administrativos.⁹² Nosotros pensamos que la fórmula matemática asociada, podría integrarse en la historia clínica electrónica, de forma que cuando el médico registre el diagnóstico, automáticamente se muestre la probabilidad de fallecimiento y con ella puedan tomarse las decisiones terapéuticas oportunas. Sin embargo, previo paso a esta sugerencia, sería necesaria la validación externa del mismo en otras áreas geográficas. En otro de los modelos que sí utilizaba datos clínicos y que podía utilizarse a pie de cama,¹⁰² únicamente se han cometido errores en el análisis estadístico (manejo de datos perdidos y utilización del análisis univariante para seleccionar predictores), pero uno de ellos podría ser crítico, pues considerar que un paciente del que no se dispone de información sobre cierto parámetro (datos perdidos), es equivalente a que dicho parámetro sea normal, no tiene mucho sentido clínico y no consideramos oportuno recomendarlo para su utilización. Finalmente, tuvimos tres modelos que, aunque alcanzaron alto riesgo de sesgo en su desarrollo,⁹⁸⁻¹⁰⁰ fueron sometidos a estudios de validación externa, lo que podría afianzar sus resultados. Si comparamos la discriminación de estos tres modelos,⁹⁸⁻¹⁰⁰ el máximo se alcanza con un área bajo la curva ROC de 0,84,¹⁰⁰ aunque todos tuvieron un valor por encima del 70%. Sin embargo, este estudio sólo ha tenido un estudio de validación externa,¹¹² mientras que existe otro trabajo con un área de 0,736,⁹⁸ que ha sido sometido a múltiples

estudios de validación.¹⁰⁸⁻¹¹¹ El tercer modelo tiene la discriminación más baja (0,712) y un único estudio de validación externa,^{99,113} aunque ha alcanzado mejores indicadores en el PROBAST. En consecuencia, si tuviéramos que recomendar alguno de estos tres modelos, nos declinaríamos preferentemente por el de Osborn,⁹⁸ pues ha sido sometido a validación externa en distintas áreas geográficas y tiene una discriminación satisfactoria. Sin embargo, los estudios de Granja y Sivayoham,^{99,100} también han sido validados externamente y tienen buena discriminación, por lo que podrían usarse con cierta prudencia. No obstante, el clínico ha de ser muy cauto en su utilización, pues los modelos presentan alto riesgo de sesgo, según PROBAST. El resto de modelos,^{89-97,101,102} no fueron, hasta donde nosotros conocemos, validados externamente, pero obtuvieron una discriminación (área bajo la curva ROC) que siempre estuvo por encima de 0,70 (máximo de 0,955, con valores oscilando generalmente alrededor de 0,75). En consecuencia, si estos modelos se validaran externamente con buenos resultados, podría considerarse también su utilización, nuevamente con cautela. En la tabla 13 se muestra a modo de resumen la discusión del estudio.

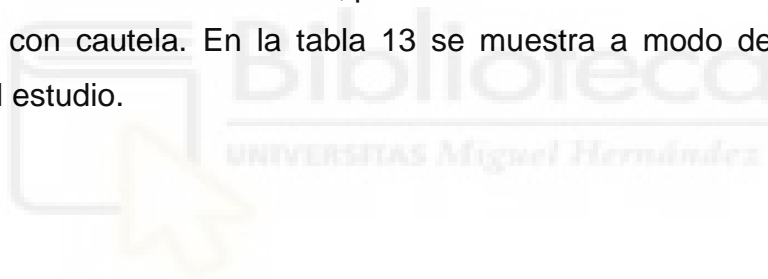


Tabla 13 Resumen apartado Discusión

Discusión											
Fortalezas	<ul style="list-style-type: none"> - Diseño sistemático. - Herramientas para el análisis desarrolladas por expertos en modelos de predicción. 										
Limitaciones	<p>Trabajos no detectados:</p> <ul style="list-style-type: none"> - No indexados en las bases de datos analizadas. - Publicados en otros idiomas. 										
Comparación con literatura	<ul style="list-style-type: none"> - Resultados similares a otras patologías. - Apartado de "Análisis" el de mayor riesgo de sesgo. 										
Implicaciones en investigación	<ul style="list-style-type: none"> - Alto riesgo de sesgo y alta preocupación respecto a su aplicabilidad en un alto porcentaje de los modelos analizados. - Se debería utilizar CHARMS y PROBAST en los analizados y en nuevos estudios. 										
Futuras líneas de investigación	<table border="1"> <tr> <td>Comparar los parámetros de discriminación y calibración entre los modelos de alto y bajo riesgo de sesgo, y valorar si es un factor determinante de la precisión.</td> <td>Revisión de estudios que realizan validación de modelos existentes, en población con sepsis.</td> </tr> </table>	Comparar los parámetros de discriminación y calibración entre los modelos de alto y bajo riesgo de sesgo, y valorar si es un factor determinante de la precisión.	Revisión de estudios que realizan validación de modelos existentes, en población con sepsis.								
Comparar los parámetros de discriminación y calibración entre los modelos de alto y bajo riesgo de sesgo, y valorar si es un factor determinante de la precisión.	Revisión de estudios que realizan validación de modelos existentes, en población con sepsis.										
Implicaciones en la práctica clínica	<table border="1"> <tr> <td> <ul style="list-style-type: none"> - Schwarkpof podría cumplir los criterios de calidad PROBAST, pero son datos administrativos y los autores no lo recomiendan. - Tabak solo errores en análisis univariante y datos perdidos, pero éste último puede ser crítico. </td> <td></td> </tr> <tr> <td> <ul style="list-style-type: none"> - Osborn es el que podríamos recomendar. </td> <td> <p>Múltiples estudios de validación externa.</p> <p>Área bajo la curva ROC 0.736</p> </td> </tr> <tr> <td> <ul style="list-style-type: none"> - Sivayoham, podría usarse con cautela. </td> <td> <p>Un estudio de validación.</p> <p>Área bajo la curva ROC 0.712</p> <p>Mejores indicadores en PROBAST</p> </td> </tr> <tr> <td> <ul style="list-style-type: none"> - Granja, podría usarse con cautela. </td> <td> <p>Un estudio de validación.</p> <p>Área bajo la curva ROC 0.84</p> </td> </tr> <tr> <td> <ul style="list-style-type: none"> - Resto de modelos podrían ser útiles, con cautela, si fueran validados. </td> <td> <p>Sin estudios de validación externa.</p> <p>Área bajo la curva ROC > 0.70</p> </td> </tr> </table>	<ul style="list-style-type: none"> - Schwarkpof podría cumplir los criterios de calidad PROBAST, pero son datos administrativos y los autores no lo recomiendan. - Tabak solo errores en análisis univariante y datos perdidos, pero éste último puede ser crítico. 		<ul style="list-style-type: none"> - Osborn es el que podríamos recomendar. 	<p>Múltiples estudios de validación externa.</p> <p>Área bajo la curva ROC 0.736</p>	<ul style="list-style-type: none"> - Sivayoham, podría usarse con cautela. 	<p>Un estudio de validación.</p> <p>Área bajo la curva ROC 0.712</p> <p>Mejores indicadores en PROBAST</p>	<ul style="list-style-type: none"> - Granja, podría usarse con cautela. 	<p>Un estudio de validación.</p> <p>Área bajo la curva ROC 0.84</p>	<ul style="list-style-type: none"> - Resto de modelos podrían ser útiles, con cautela, si fueran validados. 	<p>Sin estudios de validación externa.</p> <p>Área bajo la curva ROC > 0.70</p>
<ul style="list-style-type: none"> - Schwarkpof podría cumplir los criterios de calidad PROBAST, pero son datos administrativos y los autores no lo recomiendan. - Tabak solo errores en análisis univariante y datos perdidos, pero éste último puede ser crítico. 											
<ul style="list-style-type: none"> - Osborn es el que podríamos recomendar. 	<p>Múltiples estudios de validación externa.</p> <p>Área bajo la curva ROC 0.736</p>										
<ul style="list-style-type: none"> - Sivayoham, podría usarse con cautela. 	<p>Un estudio de validación.</p> <p>Área bajo la curva ROC 0.712</p> <p>Mejores indicadores en PROBAST</p>										
<ul style="list-style-type: none"> - Granja, podría usarse con cautela. 	<p>Un estudio de validación.</p> <p>Área bajo la curva ROC 0.84</p>										
<ul style="list-style-type: none"> - Resto de modelos podrían ser útiles, con cautela, si fueran validados. 	<p>Sin estudios de validación externa.</p> <p>Área bajo la curva ROC > 0.70</p>										

CHARMS: CHEcklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies; PROBAST: Prediction model Risk Of Bias ASsessment Tool; ROC: receiver operating characteristic



11. CONCLUSIÓN





Los modelos de predicción de mortalidad analizados en esta revisión sistemática mostraron gran variabilidad en sus diseños y metodologías. Después de aplicar PROBAST como el *gold standard* para verificar si se desarrollaron correctamente, encontramos que solamente un estudio obtuvo una puntuación positiva en la aplicabilidad junto con una valoración incierta en cuanto al riesgo de sesgo.

En el resto de estudios obtuvimos un alto riesgo de sesgo. Además, de los modelos estudiados, sólo tres fueron validados externamente, dos de ellos una vez y el otro en cuatro ocasiones.





12. REFERENCIAS





1. Zárate M, Gonzalez J, Julián-Jiménez A, et al. Estudio INFURG-SEMES: epidemiología de las infecciones atendidas en los servicios de urgencias hospitalarios y evolución durante la última década. *Emergencias* 2013; 25: 368-378. *Emergencias*. 2013;25:368-378.
2. Laguna del Estal, P. Estudio epidemiológico de las infecciones en el Área de Urgencias. *Emergencias*. 2000;12(2):80-89.
3. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*. 2016;315(8):801-810. doi:10.1001/jama.2016.0287
4. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Med*. 2017;43(3):304-377. doi:10.1007/s00134-017-4683-6
5. Evans L, Rhodes A, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock 2021. *Crit Care Med*. 2021;49(11):e1063-e1143. doi:10.1097/CCM.0000000000005337
6. Gaiieski DF, Edwards JM, Kallan MJ, Carr BG. Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med*. 2013;41(5):1167-1174. doi:10.1097/CCM.0b013e31827c09f8
7. Esteban A, Frutos-Vivar F, Ferguson ND, et al. Sepsis incidence and outcome: contrasting the intensive care unit with the hospital ward. *Crit Care Med*. 2007;35(5):1284-1289. doi:10.1097/01.CCM.0000260960.94300.DE
8. Czura CJ. "Merinoff symposium 2010: sepsis"-speaking with one voice. *Mol Med*. 2011;17(1-2):2-3. doi:10.2119/molmed.2010.00001.commentary
9. Iwashyna TJ, Ely EW, Smith DM, Langa KM. Long-term cognitive impairment and functional disability among survivors of severe sepsis. *JAMA*. 2010;304(16):1787-1794. doi:10.1001/jama.2010.1553
10. Levy MM, Fink MP, Marshall JC, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Intensive Care Med*. 2003;29(4):530-538. doi:10.1007/s00134-003-1662-x
11. Bone, RC, Balk, RA, Cerra, FB, et al. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Crit Care Med*. 1992;20(6):864-874.
12. Neira E, Malaga G. Sepsis-3 y las nuevas definiciones, ¿es tiempo de abandonar SIRS? *ACTA MEDICA PERUANA*. 2016;33:217-222. doi:10.35663/amp.2016.333.115
13. Vincent JL, Opal SM, Marshall JC, Tracey KJ. Sepsis definitions: time for change. *Lancet*. 2013;381(9868):774-775. doi:10.1016/S0140-6736(12)61815-7
14. Lai NA, Kruger P. The predictive ability of a weighted systemic inflammatory response syndrome score for microbiologically confirmed infection in hospitalised patients with suspected sepsis. *Crit Care Resusc*. 2011;13(3):146-150.

15. Jaimes F, Garcés J, Cuervo J, et al. The systemic inflammatory response syndrome (SIRS) to identify infected patients in the emergency room. *Intensive Care Med.* 2003;29(8):1368-1371. doi:10.1007/s00134-003-1874-0
16. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 1996;22(7):707-710. doi:10.1007/BF01709751
17. Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. *Crit Care Med.* 2014;42(8):1749-1755. doi:10.1097/CCM.0000000000000330
18. Ferrer R, Artigas A. Physiologic parameters as biomarkers: what can we learn from physiologic variables and variation? *Crit Care Clin.* 2011;27(2):229-240. doi:10.1016/j.ccc.2010.12.008
19. Seymour CW, Liu VX, Iwashyna TJ, 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):762-774. doi:10.1001/jama.2016.0288
20. Churpek MM, Snyder A, Han X, et al. Quick Sepsis-related Organ Failure Assessment, Systemic Inflammatory Response Syndrome, and Early Warning Scores for Detecting Clinical Deterioration in Infected Patients outside the Intensive Care Unit. *Am J Respir Crit Care Med.* 2017;195(7):906-911. doi:10.1164/rccm.201604-0854OC
21. Smith GB, Prytherch DR, Meredith P, Schmidt PE, Featherstone PI. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation.* 2013;84(4):465-470. doi:10.1016/j.resuscitation.2012.12.016
22. Subbe CP. Validation of a modified Early Warning Score in medical admissions. *QJM.* 2001;94(10):521-526. doi:10.1093/qjmed/94.10.521
23. Llopis-Roca F, López Izquierdo R, Miro O, García-Lamberechts JE, Julián Jiménez A, González Del Castillo J. [Current situation of sepsis care in Spanish emergency departments]. *Rev Esp Quimioter.* 2022;35(2):192-203. doi:10.37201/req/150.2021
24. Palencia Herrejón E, González Del Castillo J, Ramasco Rueda F, et al. Consensus document for sepsis code implementation and development in the Community of Madrid. *Rev Esp Quimioter.* 2019;32(4):400-409.
25. Yébenes JC, Lorenzo C, Esteban E, et al. Interhospital sepsis code in Catalonia (Spain): territorial model for initial care of patients with sepsis. *Medicina Intensiva (English Edition).* 2020;44(1):36-45. doi:10.1016/j.medicine.2019.09.001
26. Rascado Sedes, P, Vidal Cortés, P, Pérez del Molino Bernal, M, Pérez Rodríguez, MT, Gutiérrez Urbón, JM, Uriel Latorre, BM. Plan de asistencia a la sepsis de Galicia. Código sepsis. [Internet]. Published online 2018. <https://sergas.es/asistencia-sanitaria/sepsis>
27. Borges Sá, M, Candel González, F.J, Ferrer Roca, R, Vidal Cortés, P, Zaragoza Crespo, R, Jiménez, J. Código sepsis: documento de consenso. [Internet]. Published online 2016.

<https://www.seguridaddelpaciente.es/en/information/publicaciones/2016/codigo-sepsis-documento-de-consenso/>

28. García-López L, Grau-Cerrato S, de Frutos-Soto A, et al. Impact of the implementation of a Sepsis Code hospital protocol in antibiotic prescription and clinical outcomes in an intensive care unit. *Med Intensiva*. 2017;41(1):12-20. doi:10.1016/j.medin.2016.08.001
29. Ferreras Amez JM, Arribas Entrala B, Sarrat Torres MA, et al. [Before-after study of the effect of implementing a sepsis code for emergency departments in the community of Aragon]. *Emergencias*. 2017;29(3):154-160.
30. Levy MM, Evans LE, Rhodes A. The Surviving Sepsis Campaign Bundle: 2018 update. *Intensive Care Med*. 2018;44(6):925-928. doi:10.1007/s00134-018-5085-0
31. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(2):e1001380. doi:10.1371/journal.pmed.1001380
32. Barnaby D, Ferrick K, Kaplan DT, Shah S, Bijur P, Gallagher EJ. Heart rate variability in emergency department patients with sepsis. *Acad Emerg Med*. 2002;9(7):661-670. doi:10.1111/j.1553-2712.2002.tb02143.x
33. Vincent JL, Moreno R. Clinical review: scoring systems in the critically ill. *Crit Care*. 2010;14(2):207. doi:10.1186/cc8204
34. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPMO-III). *Crit Care Med*. 2007;35(3):827-835. doi:10.1097/01.CCM.0000257337.63529.9F
35. Moreno RP, Metnitz PGH, Almeida E, et al. SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med*. 2005;31(10):1345-1355. doi:10.1007/s00134-005-2763-5
36. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med*. 2006;34(5):1297-1310. doi:10.1097/01.CCM.0000215112.84523.F0
37. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:b375. doi:10.1136/bmj.b375
38. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013;346:e5595. doi:10.1136/bmj.e5595
39. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381. doi:10.1371/journal.pmed.1001381
40. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9:103. doi:10.1186/1741-7015-9-103

41. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63. doi:10.7326/M14-0697
42. Palazón-Bru A, Martín-Pérez F, Mares-García E, et al. A general presentation on how to carry out a CHARMS analysis for prognostic multivariate models. *Stat Med.* 2020;39(23):3207-3225. doi:10.1002/sim.8660
43. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604. doi:10.1136/bmj.b604
44. Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. *BMC Med Inform Decis Mak.* 2006;6:38. doi:10.1186/1472-6947-6-38
45. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med.* 2010;8:20. doi:10.1186/1741-7015-8-20
46. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744
47. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol.* 2010;63(8):e1-37. doi:10.1016/j.jclinepi.2010.03.004
48. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer.* 2005;93(4):387-391. doi:10.1038/sj.bjc.6602678
49. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573-577. doi:10.7326/0003-4819-147-8-200710160-00010
50. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem.* 2003;49(1):7-18. doi:10.1373/49.1.7
51. Janssens ACJW, Ioannidis JPA, Bedrosian S, et al. Strengthening the reporting of genetic risk prediction studies (GRIPS): explanation and elaboration. *Eur J Epidemiol.* 2011;26(4):313-337. doi:10.1007/s10654-011-9551-z
52. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med.* 2009;151(4):W65-94. doi:10.7326/0003-4819-151-4-200908180-00136
53. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ.* 2011;343:d5928. doi:10.1136/bmj.d5928
54. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003;3:25. doi:10.1186/1471-2288-3-25

55. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med.* 2013;158(4):280-286. doi:10.7326/0003-4819-158-4-201302190-00009
56. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med.* 2019;170(1):51-58. doi:10.7326/M18-1376
57. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med.* 2019;170(1):W1-W33. doi:10.7326/M18-1377
58. Palazón-Bru A, Mares-García E, López-Bru D, et al. A critical appraisal of the clinical applicability and risk of bias of the predictive models for mortality and recurrence in patients with oropharyngeal cancer: Systematic review. *Head Neck.* 2020;42(4):763-773. doi:10.1002/hed.26025
59. Palazón-Bru A, Mares-García E, López-Bru D, Mares-Arambul E, Gil-Guillén VF, Carbonell-Torregrosa MÁ. A systematic review of predictive models for recurrence and mortality in patients with tongue cancer. *Eur J Cancer Care (Engl).* 2019;28(6):e13157. doi:10.1111/ecc.13157
60. Ueda T, Aoyama-Ishikawa M, Nakao A, Yamada T, Usami M, Kotani J. A simple scoring system based on neutrophil count in sepsis patients. *Med Hypotheses.* 2014;82(3):382-386. doi:10.1016/j.mehy.2014.01.007
61. Mikacenic C, Price BL, Harju-Baker S, et al. A Two-Biomarker Model Predicts Mortality in the Critically Ill with Sepsis. *Am J Respir Crit Care Med.* 2017;196(8):1004-1011. doi:10.1164/rccm.201611-2307OC
62. Moreno RP, Metnitz B, Adler L, et al. Sepsis mortality prediction based on predisposition, infection and response. *Intensive Care Med.* 2008;34(3):496-504. doi:10.1007/s00134-007-0943-1
63. Howell MD, Talmor D, Schuetz P, Hunziker S, Jones AE, Shapiro NI. Proof of principle: the predisposition, infection, response, organ failure sepsis staging system. *Crit Care Med.* 2011;39(2):322-327. doi:10.1097/CCM.0b013e3182037a8e
64. Shapiro NI, Wolfe RE, Moore RB, Smith E, Burdick E, Bates DW. Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule. *Crit Care Med.* 2003;31(3):670-675. doi:10.1097/01.CCM.0000054867.01688.D1
65. Chen KF, Liu SH, Li CH, et al. Development and validation of a parsimonious and pragmatic CHARM score to predict mortality in patients with suspected sepsis. *Am J Emerg Med.* 2017;35(4):640-646. doi:10.1016/j.ajem.2016.10.075
66. Fortis S, O'Shea AMJ, Beck BF, et al. An automated computerized critical illness severity scoring system derived from APACHE III: modified APACHE. *J Crit Care.* 2018;48:237-242. doi:10.1016/j.jcrc.2018.09.005
67. Jaimes F, Farbiarz J, Alvarez D, Martínez C. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. *Crit Care.* 2005;9(2):R150-156. doi:10.1186/cc3054

68. Shapiro NI, Howell MD, Talmor D, Donnino M, Ngo L, Bates DW. Mortality in Emergency Department Sepsis (MEDS) score predicts 1-year mortality. *Crit Care Med.* 2007;35(1):192-198. doi:10.1097/01.CCM.0000251508.12555.3E
69. Sivayoham N, Blake LA, Tharimooopantavida SE, et al. The REDS score: a new scoring system to risk-stratify emergency department suspected sepsis: a derivation and validation study. *BMJ Open.* 2019;9(8):e030922. doi:10.1136/bmjopen-2019-030922
70. Ward L, Paul M, Andreassen S. Automatic learning of mortality in a CPN model of the systemic inflammatory response syndrome. *Math Biosci.* 2017;284:12-20. doi:10.1016/j.mbs.2016.11.004
71. Zhang Z, Hong Y. Development of a novel score for the prediction of hospital mortality in patients with severe sepsis: the use of electronic healthcare records with LASSO regression. *Oncotarget.* 2017;8(30):49637-49645. doi:10.18632/oncotarget.17870
72. Garcia-Simon M, Morales JM, Modesto-Alapont V, et al. Prognosis Biomarkers of Severe Sepsis and Septic Shock by 1H NMR Urine Metabolomics in the Intensive Care Unit. *PLoS One.* 2015;10(11):e0140993. doi:10.1371/journal.pone.0140993
73. Rubulotta F, Marshall JC, Ramsay G, Nelson D, Levy M, Williams M. Predisposition, insult/infection, response, and organ dysfunction: A new model for staging severe sepsis. *Crit Care Med.* 2009;37(4):1329-1335. doi:10.1097/CCM.0b013e31819d5db1
74. Santos TM, Franci D, Coutinho CMG, et al. A simplified ultrasound-based edema score to assess lung injury and clinical severity in septic patients. *Am J Emerg Med.* 2013;31(12):1656-1660. doi:10.1016/j.ajem.2013.08.053
75. Saka G, Kreke JE, Schaefer AJ, Chang CCH, Roberts MS, Angus DC. Use of dynamic microsimulation to predict disease progression in patients with pneumonia-related sepsis. *Crit Care.* 2007;11(3):R65. doi:10.1186/cc5942
76. Dumas G, Lavillegrand JR, Joffre J, et al. Mottling score is a strong predictor of 14-day mortality in septic patients whatever vasopressor doses and other tissue perfusion parameters. *Crit Care.* 2019;23(1):211. doi:10.1186/s13054-019-2496-4
77. Asiimwe SB, Abdallah A, Ssekitoleso R. A simple prognostic index based on admission vital signs data among patients with sepsis in a resource-limited setting. *Crit Care.* 2015;19:86. doi:10.1186/s13054-015-0826-8
78. Baumgartner JD, Büla C, Vaney C, Wu MM, Eggimann P, Perret C. A novel score for predicting the mortality of septic shock patients. *Crit Care Med.* 1992;20(7):953-960. doi:10.1097/00003246-199207000-00010
79. Guillén J, Liu J, Furr M, et al. Predictive models for severe sepsis in adult ICU patients. *2015 Systems and Information Engineering Design Symposium.* Published online 2015. doi:10.1109/SIEDS.2015.7116970
80. Hebert PC, Drummond AJ, Singer J, Bernard GR, Russell JA. A simple multiple system organ failure scoring system predicts mortality of patients who have sepsis syndrome. *Chest.* 1993;104(1):230-235. doi:10.1378/chest.104.1.230

81. Slotman GJ. Prospectively validated prediction of physiologic variables and organ failure in septic patients: The Systemic Mediator Associated Response Test (SMART). *Crit Care Med.* 2002;30(5):1035-1045. doi:10.1097/00003246-200205000-00014
82. Lagu T, Rothberg M, Nathanson, BH, Steingrub, JS, Lindenauer, PK. Incorporating initial treatments improves performance of a mortality prediction model for patients with sepsis. *Pharmacoepidemiol Drug Saf.* 2012;21 Suppl 2:44-52. doi:10.1002/pds.3229
83. Sivayoham N, Holmes P, Cecconi M, Rhodes A. Internal emergency department validation of the simplified MISSED score. *Eur J Emerg Med.* 2015;22(5):321-326. doi:10.1097/MEJ.0000000000000176
84. Lissalde-Lavigne G, Combescure C, Muller L, et al. Simple coagulation tests improve survival prediction in patients with septic shock. *J Thromb Haemost.* 2008;6(4):645-653. doi:10.1111/j.1538-7836.2008.02895.x
85. Kim YC, Song JE, Kim EJ, et al. A Simple Scoring System Using the Red Blood Cell Distribution Width, Delta Neutrophil Index, and Platelet Count to Predict Mortality in Patients With Severe Sepsis and Septic Shock. *J Intensive Care Med.* 2019;34(2):133-139. doi:10.1177/0885066618787448
86. Perl TM, Dvorak L, Hwang T, Wenzel RP. Long-term survival and function after suspected gram-negative sepsis. *JAMA.* 1995;274(4):338-345.
87. West TE, Wikraiphat C, Tandhavanant S, et al. Patient Characteristics, Management, and Predictors of Outcome from Severe Community-Onset Staphylococcal Sepsis in Northeast Thailand: A Prospective Multicenter Study. *Am J Trop Med Hyg.* 2017;96(5):1042-1049. doi:10.4269/ajtmh.16-0606
88. Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. *Acad Emerg Med.* 2016;23(3):269-278. doi:10.1111/acem.12876
89. García-Lamberechts EJ, Martín-Sánchez FJ, Julián-Jiménez A, et al. Infection and systemic inflammatory response syndrome in older patients in the emergency department: a 30-day risk model. *Emergencias.* 2018;30(4):241-246.
90. Jiang W, Zhong W, Deng Y, et al. Evaluation of a combination "lymphocyte apoptosis model" to predict survival of sepsis patients in an intensive care unit. *BMC Anesthesiol.* 2018;18(1):89. doi:10.1186/s12871-018-0535-3
91. Samsudin MI, Liu N, Prabhakar SM, et al. A novel heart rate variability based risk prediction model for septic patients presenting to the emergency department. *Medicine (Baltimore).* 2018;97(23):e10866. doi:10.1097/MD.00000000000010866
92. Schwarzkopf D, Fleischmann-Struzek C, Rüdell H, Reinhart K, Thomas-Rüdell DO. A risk-model for hospital mortality among patients with severe sepsis or septic shock based on German national administrative claims data. *PLoS One.* 2018;13(3):e0194371. doi:10.1371/journal.pone.0194371
93. Bewersdorf JP, Hautmann O, Kofink D, Abdul Khalil A, Zainal Abidin I, Loch A. The SPEED (sepsis patient evaluation in the emergency department) score: a risk stratification and

- outcome prediction tool. *Eur J Emerg Med.* 2017;24(3):170-175. doi:10.1097/MEJ.0000000000000344
94. Fang WF, Douglas IS, Chen YM, et al. Development and validation of immune dysfunction score to predict 28-day mortality of sepsis patients. *PLoS One.* 2017;12(10):e0187088. doi:10.1371/journal.pone.0187088
 95. Ford DW, Goodwin AJ, Simpson AN, Johnson E, Nadig N, Simpson KN. A Severe Sepsis Mortality Prediction Model and Score for Use With Administrative Data. *Crit Care Med.* 2016;44(2):319-327. doi:10.1097/CCM.0000000000001392
 96. Seo MH, Choa M, You JS, et al. Hypoalbuminemia, Low Base Excess Values, and Tachypnea Predict 28-Day Mortality in Severe Sepsis and Septic Shock Patients in the Emergency Department. *Yonsei Med J.* 2016;57(6):1361-1369. doi:10.3349/ymj.2016.57.6.1361
 97. Mohan A, Shrestha P, Guleria R, Pandey RM, Wig N. Development of a mortality prediction formula due to sepsis/severe sepsis in a medical intensive care unit. *Lung India.* 2015;32(4):313-319. doi:10.4103/0970-2113.159533
 98. Osborn TM, Phillips G, Lemeshow S, et al. Sepsis severity score: an internationally derived scoring system from the surviving sepsis campaign database*. *Crit Care Med.* 2014;42(9):1969-1976. doi:10.1097/CCM.0000000000000416
 99. Sivayoham N, Rhodes A, Cecconi M. The MISSED score, a new scoring system to predict Mortality In Severe Sepsis in the Emergency Department: a derivation and validation study. *Eur J Emerg Med.* 2014;21(1):30-36. doi:10.1097/MEJ.0b013e328364a8d4
 100. Granja C, Póvoa P, Lobo C, Teixeira-Pinto A, Carneiro A, Costa-Pereira A. The predisposition, infection, response and organ failure (Piro) sepsis classification system: results of hospital mortality using a novel concept and methodological approach. *PLoS One.* 2013;8(1):e53885. doi:10.1371/journal.pone.0053885
 101. Lagu T, Lindenauer PK, Rothberg MB, et al. Development and validation of a model that uses enhanced administrative data to predict mortality in patients with sepsis. *Crit Care Med.* 2011;39(11):2425-2430. doi:10.1097/CCM.0b013e31822572e3
 102. Tabak YP, Johannes RS, Silber JH. Using automated clinical data for risk adjustment: development and validation of six disease-specific mortality predictive models for pay-for-performance. *Med Care.* 2007;45(8):789-805. doi:10.1097/MLR.0b013e31803d3b41
 103. Du M, Haag D, Song Y, Lynch J, Mittinty M. Examining Bias and Reporting in Oral Health Prediction Modeling Studies. *J Dent Res.* 2020;99(4):374-387. doi:10.1177/0022034520903725
 104. Di Tanna GL, Wirtz H, Burrows KL, Globe G. Evaluating risk prediction models for adults with heart failure: A systematic literature review. *PLoS One.* 2020;15(1):e0224135. doi:10.1371/journal.pone.0224135
 105. Bellou V, Belbasis L, Konstantinidis A, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: Systematic review and critical appraisal. *BMJ.* 2019;367:l5358. doi:10.1136/bmj.l5358

106. Carrillo-Larco RM, Aparcana-Granda DJ, Mejia JR, Barengo NC, Bernabe-Ortiz A. Risk scores for type 2 diabetes mellitus in Latin America: a systematic review of population-based studies. *Diabet Med.* 2019;36(12):1573-1584. doi:10.1111/dme.14114
107. Carrillo-Larco RM, Altez-Fernandez C, Pacheco-Barrios N, et al. Cardiovascular Disease Prognostic Models in Latin America and the Caribbean: A Systematic Review. *Glob Heart.* 2019;14(1):81-93. doi:10.1016/j.gheart.2019.03.001
108. Wełna M, Adamik B, Goździk W, Kübler A. External validation of the sepsis severity score. *Int J Immunopathol Pharmacol.* 2020;34:2058738420936386. doi:10.1177/2058738420936386
109. Sathaporn N, Khwannimit B. Validation the performance of New York Sepsis Severity Score compared with Sepsis Severity Score in predicting hospital mortality among sepsis patients. *J Crit Care.* 2019;53:155-161. doi:10.1016/j.jcrc.2019.06.017
110. Khwannimit B, Bhurayanontachai R, Vattanavanit V. Validation of the Sepsis Severity Score Compared with Updated Severity Scores in Predicting Hospital Mortality in Sepsis Patients. *Shock.* 2017;47(6):720-725. doi:10.1097/SHK.0000000000000818
111. Williams JM, Greenslade JH, Chu K, Brown AFT, Lipman J. Severity Scores in Emergency Department Patients With Presumed Infection: A Prospective Validation Study. *Crit Care Med.* 2016;44(3):539-547. doi:10.1097/CCM.0000000000001427
112. Caramello V, Macciotta A, Beux V, De Salve AV, Ricceri F, Boccuzzi A. Validation of the Predisposition Infection Response Organ (PIRO) dysfunction score for the prognostic stratification of patients with sepsis in the Emergency Department. *Med Intensiva.* Published online June 23, 2020. doi:10.1016/j.medin.2020.04.022
113. Ryoo SM, Ahn S, Kim WY, Lim KS. External validation of the MISSED score to predict mortality in patients with severe sepsis and septic shock in the emergency department. *Eur J Emerg Med.* 2015;22(5):327-330. doi:10.1097/MEJ.0000000000000156



13. ANEXO





TITLE PAGE:

Title: A critical appraisal of the prognostic predictive models for patients with sepsis: which model can be applied in clinical practice?

Running title: Prognostic models in sepsis.

Authors: Concepción Beneyto-Ripoll, MD¹, Antonio Palazón-Bru, PhD², Patricia Llópez-Espinós, MD³, Ana María Martínez-Díaz, MD⁴, Vicente Francisco Gil-Guillén, MD, PhD^{2*}, María de los Ángeles Carbonell-Torregrosa, MD, PhD^{2,3*}.

Institutions:

1. Emergency Services, General Hospital of Almansa, Almansa, Albacete, Spain.
2. Department of Clinical Medicine, Miguel Hernández University, San Juan de Alicante, Alicante, Spain.
3. Emergency Services, General University Hospital of Elda, Elda, Alicante, Spain.
4. Emergency Services, University Hospital of Puerta del Mar, Cádiz, Cádiz, Spain.

*, equal contribution

Corresponding author: Prof. Antonio Palazón-Bru, PhD. Department of Clinical Medicine, Miguel Hernández University, Ctra Valencia-Alicante S/N, 03550 San Juan de Alicante, Alicante, Spain. Phone: +34965919449. Fax: +34965919450. E-mail: antonio.pb23@gmail.com.

Acknowledgments: The authors thank Maria Repice and Ian Johnstone for their help in the English version of the text. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.



ABSTRACT

Background: Sepsis is associated with high mortality and predictive models can help in clinical decision making. The objective of this study was to carry out a systematic review of these models.

Methods: In 2019, we conducted a systematic review in MEDLINE and EMBASE (CDR4201811121:PROSPERO) of articles that developed predictive models for mortality in septic patients (inclusion criteria). We followed the CHARMS recommendations (Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies), extracting the information from its 11 domains (Source of data, Participants, etc.). We determined the risk of bias and applicability (participants, outcome, predictors and analysis) through PROBAST (Prediction model Risk Of Bias ASsessment Tool).

Results: A total of 14 studies were included. In the CHARMS extraction, the models found showed great variability in its 11 domains. Regarding the PROBAST checklist, only one article had an unclear risk of bias as it did not indicate how missing data were handled, while the others all had a high risk of bias. This was mainly due to the statistical analysis (inadequate sample size, handling of continuous predictors, missing data and selection of predictors), since 13 studies had a high risk of bias. Applicability was satisfactory in six articles. Most of the models integrate predictors from routine clinical practice. Discrimination and calibration were assessed for almost all the models, with the area under the ROC curve ranging from 0.59 to 0.955 and no lack of calibration. Only three models were externally validated and their maximum discrimination values in the derivation were from 0.712 and 0.84. One of them (Osborn) had undergone multiple validation studies.

Discussion: Despite most of the studies showing a high risk of bias, we very cautiously recommend applying the Osborn model, as this has been externally validated various times.

KEYWORDS: Death; Mortality; Sepsis/mortality; Shock, Septic/mortality; Models, Statistical.

REVIEW CRITERIA

We conducted a systematic review to evaluate the quality of prognostic models in sepsis and to determine the suitability of integrating these models into clinical practice, thereby improving decision-making. Each predictive model was assessed with the PROBAST checklist by two independent reviewers.

MESSAGE FOR THE CLINIC

Categorizing the severity of septic patients is a fundamental pillar for the management of these patients, and predictive models are effective tools for this purpose. Predictive models have proliferated in recent years and clinical practice guidelines further recommend their use. However, the models must be reliable and scientifically supported. In light of our results, caution must be exercised when using these predictive models for bedside use in sepsis.



INTRODUCTION

The incidence of sepsis has been increasing in recent years (from 300 to 1,031 cases per 100,000 population) due to various factors, such as increased life expectancy, increased performance of invasive techniques or states of drug-induced immunosuppression, all of which contribute to an increased susceptibility to infection.¹ Sepsis affects millions of people worldwide each year,² and it has been established as one of the leading causes of death.³ However, the early, targeted application of a range of diagnostic-therapeutic measures significantly improves survival in patients with sepsis.¹

Categorizing the severity of septic patients is a fundamental pillar for the early identification and structured management of patients at increased risk of mortality. Tools exist to accomplish this task, such as risk prediction models or prognostic scales, on which medical professionals rely to make clinical decisions.⁴ A prognostic model is a formal combination of multiple predictors from which the risks of a specific endpoint can be calculated for individual patients,⁵ helping us to make decisions to improve clinical outcomes in our patients.^{6,7} In fact, in recent years risk prediction models in medicine have proliferated, and clinical practice guidelines increasingly recommend their use,⁸ even though they may present limitations for use, like being difficult to apply or taking up a lot of time.

Prediction models must be reliable and supported by scientific evidence.⁴ We must therefore carry out a critical analysis of how they have been constructed/validated and of their potential use in clinical practice. In 2014, the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist was introduced, providing guidelines for the systematic review of predictive models for a specific disease. It includes seven elements to guide the objective of the review, the search strategy and the inclusion and exclusion criteria of the study as well as the information to be extracted and assessed for each model included in the review, classified into 11 domains (source of data, participants, outcome to be predicted, candidate predictors, sample size, missing data, model development, model performance, model evaluation, results, and interpretation

and discussion). The CHARMS checklist indicates how each domain should be addressed, based on the scientific literature.⁹

In 2019, complementing the CHARMS checklist,⁹ the Prediction model Risk Of Bias ASsessment Tool (PROBAST) statement emerged, which assesses the risk of bias and the applicability of predictive models.^{10,11} This would be equivalent to the Cochrane Risk of Bias Tool for assessing the quality of clinical trials contained in a systematic review,¹² but in CHARMS reviews of predictive models.⁹ PROBAST assesses the risk of bias across four domains (participants, predictors, outcome, and analysis) and applicability across three of these (participants, predictors and outcome). Each domain is rated as having a low, high or unclear risk of bias/concern regarding applicability. Finally, an overall assessment is given, both for the risk of bias and for applicability. Each domain is scored objectively under the PROBAST recommendations, which, like CHARMS,⁹ are based on the scientific literature.^{10,11}

Sepsis is associated with a high mortality rate (~25%),² and it is essential to make an early determination of the severity of the disease for the early implementation of diagnostic-therapeutic measures adjusted to this prognostic risk. However, as far as we know, no systematic review of prediction models has been carried out with the application of PROBAST on prognostic models for this disease. Accordingly, the aim of this study was to perform a systematic review with these characteristics, providing a critical document for determining the suitability of integrating existing prognostic models for septic patients in clinical practice.

METHODS

Protocol and registration

This review was registered in the PROSPERO database (CDR4201811121), where the methods on which it would be based were specified a priori. In addition, this review is based on the CHARMS and PROBAST statements.⁹⁻¹¹

Eligibility criteria

We conducted a systematic review of scientific articles that developed a prediction model (at least two predictors) of mortality (prognostic model) in which the authors used their own data set in patients with sepsis. Internal or external validation of the model could also have been carried out in the publication. Internal validation is what is done on the data set used to develop the model and external validation is that done on any other data set. Models involving samples of patients with sepsis in general, i.e. not caused by a specific pathogen, were considered. The definition of sepsis used in our study was that in effect at the time each of the models was developed. Our aim was to obtain models that could be used at the time of diagnosis or admission of the patient for sepsis (baseline condition), to predict all-cause mortality (outcome to be predicted), regardless of the time of death from baseline. The designs that correspond with all these characteristics are cohort studies, whether prospective or retrospective, since we are going to predict a future event (mortality).⁹

We excluded review articles or articles describing the development of a model in which the authors did not use their own data (e.g. simulated data), conference summaries, studies that analyzed the value of new predictors in existing models, articles that only performed an external validation of a prediction model already published or that did not develop a multivariate risk model, studies conducted in a pediatric or surgical population, articles that referred to bacteremia or another infectious entity that did not define sepsis, or those that referred to a specific microorganism that caused the infection. Articles that applied Machine Learning techniques were also excluded, except for those analytic techniques which used logistic regression models. The other kind of models are not suitable for the CHARMS statement and PROBAST checklist.

Information sources

The MEDLINE (through Pubmed), Scopus and EMBASE databases were analyzed from inception to September 17, 2019.

Search

The databases were examined using keywords referring to several terms: prediction models (predictive model, prediction model, risk score, scoring system, points system and nomogram), type of patients (sepsis) and outcome (mortality, death and prognosis). These keywords were searched in the title, abstract or thesaurus of each database analyzed (MESH and Emtree). Articles were restricted to those published in English or Spanish with an abstract (languages that the authors of this review understand). We analyzed the references of all articles that could be relevant to our research question. The search equations are given in Appendix E1.

Two blinded reviewers independently analyzed titles and abstracts to determine which scientific articles could potentially be included. Results were then pooled and disagreements were resolved by consensus. The same procedure was applied to the full text of articles associated with these titles and abstracts, as well as to the references of included articles. These reviewers were experts in prediction models and/or septic disease. No hand searching was done to analyze the gray literature as the topic is complex and specific to scientific journals.

Data extraction

Of the articles finally included in the review, the same two blinded reviewers independently extracted the information required by the CHARMS checklist.⁹ This was grouped into 11 domains: 1) Source of data (study design); 2) Participants (description, dates and treatments); 3) Outcome to be predicted (definition of the outcome to be predicted and period of onset); 4) Candidate predictors (number and type, form and time of measurement, and handling of continuous predictors); 5) Sample size (number of participants, number of events and the event-per-variable ratio); 6) Missing data (number of participants with missing data overall and for each predictor, and handling of missing data); 7) Model development (type and methods used for the selection of predictors); 8) Model performance (calibration and discrimination); 9) Model evaluation (method used); 10) Results (form of presentation

of the final model (nomogram, points system, mathematical formula, etc.), weight of the different variables, and comparison of development and external validation samples when appropriate); and 11) Interpretation and discussion (comparison with similar studies, generalization of the model, strengths and limitations).⁹ All this information is presented in tabular form to make it easier to see all the characteristics of the models.^{13,14}

Quality assessment

After having extracted in duplicate the information indicated in CHARMS,⁹ the PROBAST tool was applied, which assesses both the risk of bias and the applicability of predictive models.^{10,11} The risk of bias is divided into four domains: participants, predictors, outcome and analysis; and the applicability into three domains (participants, predictors and outcome). The *Participants* domain refers to the selection of the study sample, *Predictors* evaluates the candidate predictors in the model, *Outcome* analyzes the variable we are predicting, and *Analysis* addresses the statistical methods. The PROBAST statement indicates in an objective way how to assess each of these domains. The seven domains can be categorized into high, low or unclear, either for risk of bias or concern regarding applicability. After assessing the domains independently, an overall assessment is issued for the two blocks, using the "the worst score counts" principle, meaning that a negative item in a domain renders the overall assessment negative, and therefore its use is discouraged. For example, if in risk of bias we obtained a low level for all domains except for *Analysis*, which was high, the overall assessment of the risk of bias will be high (the worst score). The risk of bias determines whether systematic errors were made in the study, while the applicability indicates whether the model fits the research question of the systematic review,^{10,11} which in this case was the determination of the risk of death in a patient with sepsis.

Statistical analysis

A descriptive analysis of both CHARMS and PROBAST domains was performed,⁹⁻¹¹ as it does not make sense to perform a meta-analysis in this type of study. Furthermore, so as to compare our results with other similar studies, we searched systematic reviews using the PROBAST checklist in other diseases, through the analysis of their citations (Scopus), obtaining the percentage for low risk of bias and low concern regarding applicability for all the domains. We also used Scopus to analyze the articles that had been referenced in the studies included in this review to determine whether the studies had been validated externally.

RESULTS

Figure 1 shows the scheme of the systematic review. A total of 3,397 articles were screened by title and abstract in the databases analyzed (485 in MEDLINE, 1,254 in Scopus and 1,658 in EMBASE), and 44 articles were studied for full text analysis.¹⁵⁻⁵⁸ Of these, 12 were excluded because not all the patients were diagnosed with sepsis,¹⁵⁻²⁶ seven did not develop a predictive model,²⁷⁻³³ in four the outcome to be predicted was not mortality,³⁴⁻³⁷ in another four they did not use their own data,³⁸⁻⁴¹ two evaluated only one specific microorganism and not sepsis in general,^{42,43} and one applied Machine Learning techniques,⁴⁴ which were not considered in this review. Finally, 14 met all the inclusion criteria and none of the exclusion criteria,⁴⁵⁻⁵⁸ from which all the information was extracted according to CHARMS, and the risk of bias and applicability was assessed according to PROBAST.

Tables E1-5 provide the 11 CHARMS domains in full detail. Upon a descriptive analysis, half of the studies were designed prospectively and half were designed retrospectively (*Source of data*). With regard to *Participants*, we note that these were generally adult patients with sepsis or severe sepsis, often originating in the emergency department or intensive care unit. In addition, we would like to point out that there were nine studies that performed external validation on another data set.⁴⁸⁻

^{52,54,55,57,58} Treatment was assessed as a candidate predictor in five studies.^{51,53,54,56,57} Data collection dates ranged between 2000 and 2016. *Outcome to be predicted* was always all-cause mortality, indicating onset of mortality in the hospitalization period or admission to the intensive care unit. Some articles also indicated the prediction time, with a range of 28-30 days.^{45-47,49,50,52}

Candidate Predictors showed great variability, and were generally classified into demographics, comorbidities, prognostic scores (predictive models of severity in Intensive Care Units), vital signs, laboratory data, and infection characteristics/treatment. Most of these predictors were obtained at the time of admission, but only seven articles included all the variables necessary to apply the model at the time of diagnosis (baseline), that is all the information about the predictors was available at baseline.^{45,47-50,55,58} Regarding the management of continuous predictors, most used categorizations.^{45,49-51,53-56,58} In the *Sample size* domain, a range of events per variable was observed that oscillated between 0.57 and 572.28, although there were three articles in which this could not be calculated, either because the number of events or the predictors were unknown.^{51,54,58} In the articles that validated the model constructed on another data set,^{48-52,54,55,57,58} the number of events oscillated between nine and 56,579, although there were two articles that did not provide this information in the publication.^{54,58}

In the *Missing data* domain, most articles did not provide this information.^{45-48,51-53,56,57} Three used complete-case analysis,^{49,50,55} and only one used multiple imputation, but with only one predictor.⁵⁴ With regard to *Model development*, all the articles estimated binary logistic regression models,⁴⁵⁻⁵⁸ and nine authors used univariate analysis to select predictors for inclusion in the multivariate model.^{45-47,49,52,54-56,58}

In *Model performance*, we emphasize that in all the models discrimination was calculated using the area under the receiver operating characteristic curve.⁴⁵⁻⁵⁸ Calibration was carried out mainly by the Hosmer-Lemeshow test and/or graphically, although in three models this aspect was not addressed.^{46,47,50} In *Model evaluation*, only three models applied resampling techniques

(bootstrapping),^{45,54,58} and in another two, although this type of technique was not applied, overfitting was taken into account.^{48,49} The values obtained for discrimination and calibration for each model are shown in Table 1, where we see that the area under the receiver operating characteristic curve ranged from 0.59 to 0.955 and no lack of calibration was found.

In *Results* it is striking that only two groups of authors obtained all the results that had been proposed in the methodology,^{46,54} with the absence of the intercept of the regression model being the most common omission.^{45,47,49,50,52,53,55-58} Another important point is the way in which the model is presented, with the most common option being a scoring system (n=7),^{45,49-51,53,54,55} followed by a mathematical formula (regression coefficients, n=6),^{46-48,56-58} and a nomogram (n=1).⁵² Finally, in *Interpretation and discussion*, we note that the majority of the articles (13/14) indicated in their discussion that caution should be exercised in the use of the model (exploratory results), but one of these directly recommended the implementation of the model in routine clinical practice.⁵⁸ A total of 12 articles discussed both the predictors found in the models and the comparison of their model with previously published models.^{45-49,52-58} The strengths and limitations were assessed by all articles.⁴⁵⁻⁵⁸ Finally, two articles did not discuss the generalization of the model to other geographical areas^{54,58}

The PROBAST analysis is reported in Table 2 and in more detail in Appendix E2. Here we find the domains for both *risk of bias* and *applicability*. Regarding the domains common to these two aspects (*participants, predictors* and *outcome*), the same assessment was obtained for both *risk of bias* and *applicability*. In *Participants* a positive score was found for 10 studies (71.4%),^{45,48-52,54-56,58} and in one there was a lack of information to determine the item.⁵⁷ For *Predictors* half of the articles were of satisfactory quality,^{45,47-49,52,55,58} and in three this was not clear.^{46,50,57} In *Outcome* the maximum score was found in 12 articles (85.7%),^{45,47-49,51-58} remaining unclear in the rest.^{46,50} The overall assessment of these three items led to a total of six studies with low concern regarding applicability.^{45,48,49,52,55,58}

The domain of *Analysis*, which is only present in the assessment of *risk of bias*, resulted in 13 studies with negative values,^{45-47,49-58} and one of these could not be assessed.⁴⁸ We must take into account

that this domain has a total of nine questions, and if one of these is assessed negatively, the domain is classified as having a high risk of bias ("the worst score counts"). Among the 14 articles included, only one had eight questions assessed positively (the article failed to indicate how it handled missing data, unclear),⁴⁸ another had seven questions assessed satisfactorily while the other two were negative (*missing data* from combining the missing data with the reference category and using univariate analysis to select the model predictors),⁵⁸ and three articles had six positive questions: 1) Ford et al., 2016: used categorizations in the continuous predictors, did not consider overfitting in the internal validation and *missing data* could not be assessed⁵¹; 2) Osborn et al., 2014: did not adequately handle continuous predictors, used univariate analysis to select model predictors, and we could not assess *missing data*⁵⁴; and 3) Lagu et al., 2011: did not take into account overfitting, and there were two items for which we had no information for their assessment, the first being *missing data* and the second the questionable use of all patients to develop the model since when making exclusions after starting the follow-up, this was unclear.⁵⁷

Of note in the studies of other diseases consulted^{13,59-63} was that analysis was the domain that presented the greatest risk of bias, as in various reviews either no or very few studies among those included did this correctly. The other domains, however, both for risk of bias and applicability, were quite satisfactory, surpassing 70% in almost all the reviews.

Of the studies found in this systematic review, only three predictive models had been subjected to external validation: those led by Osborn,⁵⁴ Sivayoham,⁵⁵ and Granja.⁵⁶ The first had four external validation studies,⁶⁴⁻⁶⁷ and the others just one (Table 2).^{68,69}

DISCUSSION

Strengths and limitations

The main strength of this study is that using a systematic design in which two people independently review all the information (around 2,500 scientific articles) and evaluate the data using tools developed by experts in predictive models⁹⁻¹¹ reduces the risk of systematic errors.

It is possible that we could have failed to detect studies that did develop a prediction model with the predetermined selection criteria, but this was minimized as the review was performed systematically.

It is also possible that other articles not indexed in MEDLINE, EMBASE or Scopus exist in other databases. Moreover, there could be publications that are indexed in these databases, but in other languages that the authors do not understand. Nevertheless, the vast majority of the scientific literature is published in English and, in fact, all the articles included in our review are in English, except for one that is available in both Spanish and English.⁴⁵ Finally, this type of review is a methodological critique of predictive models; it does not include the models themselves, although many details about these models are mentioned. Thus, if readers wish to delve deeper into the models they should refer to the original publication of the model, where the authors will have indicated how to apply it for a real patient.

Comparison with the existing literature

When comparing our results (risk of bias and applicability), we had to analyze systematic reviews with similar methodologies that used PROBAST, and we found great variability in patient types and outcomes.^{13,59-63} The percentages of low risk and low concern regarding applicability of all these studies are indicated in Table 3. Here we can see that the results obtained in sepsis for all the domains, except for *Predictors*, are in the range of the other studies. This could be due to the fact that the other models for sepsis (Table E1-5) have laboratory parameters and treatment information that may not be available at the time of making the prediction. However, this is difficult to compare, as they are very different diseases and predictors.

Implications to research and clinical practice

Given the results found in our review, we believe there are two points of improvement regarding the research, since high risk of bias and high concern regarding applicability was found in an important percentage of the models. First, when developing a new predictive model, it should be based on both the CHARMS and the PROBAST statements⁹⁻¹¹ to allow the model to be implemented in routine clinical practice, although it is important to note that the decision to implement a risk prediction model into clinical practice is not based purely on reporting items on a checklist or risk of bias. There are other important aspects, like its accuracy (discrimination and calibration) and ease of use, in addition to it having been validated externally several times. The second point requiring improvement is the use of the databases of the articles analyzed (Tables E1-5), correctly selecting patients and predictors, taking into account the aspects of the definition of the outcome and applying the recommended statistical techniques. In this way, without spending a great deal of effort collecting data from a large number of patients again, we can correct the deficiencies found.

An interesting idea would be to compare the parameters of discrimination and calibration between the studies with a high risk of bias and those with a low risk of bias, examining whether this factor might alter the accuracy of the model. As we had no study with a low risk of bias we were unable to carry out this comparison, though it would be interesting to do in another review like this one.

Another future line of useful research for clinical practice would be to undertake a review of predictive models but assessing only the external validations undertaken on patients with sepsis. This is a different review from that undertaken here as it corresponds to subtype 3 of models evaluated in the CHARMS declaration: "External model validation studies with or without model updating".⁹ The studies in this new review could include validations of models developed in patients with sepsis, like those included in our study, or models undertaken in other types of populations but validated in patients with sepsis. This new review would complement that developed in the present paper.

With respect to clinical practice, we must be very cautious, given that only a single model met the quality standards of PROBAST.⁴⁸ However, the authors themselves indicate that the model developed

should not be applied to clinical practice, since it uses administrative data.⁴⁸ We believe that the associated mathematical formula could be integrated into the electronic medical record, so that when the physician records the diagnosis, the probability of death is automatically shown and with it the appropriate therapeutic decisions can be made. However, prior to proceeding with this suggestion, external validation is needed in other geographical areas. In another model that did use clinical data and which could be used at the bedside,⁵⁸ errors were made only in the statistical analysis (missing data and use of a single variable to select predictors), though one of these could be critical, since considering that a patient for whom there is no information regarding a certain parameter (missing data) is equivalent to that parameter being normal makes little clinical sense and we do not consider it appropriate to recommend use of this model. Finally, there were three models which, although having a high risk of bias in their development,⁵⁴⁻⁵⁶ were nevertheless submitted to external validation, which could strengthen their results. Comparison of the discrimination of these three models⁵⁴⁻⁵⁶ showed that the maximum area under the ROC curve was 0.84,⁵⁶ although in all cases this value was above 70%. However, this study had only undergone one external validation process,⁶⁸ whereas another study with an area of 0.736⁵⁴ had been submitted to various validation studies.⁶⁴⁻⁶⁷ The third model had the lowest discrimination (0.712) and just one external validation study,^{55,69} although it attained better indicators in PROBAST. Consequently, if we had to recommend one of these three studies we would opt for that of Osborn,⁵⁴ as it has undergone external validation in different geographical areas and has a satisfactory discrimination. Nevertheless, the studies of Granja and Sivayoham^{55,56} were also validated externally with good discrimination, so that these too could also be used cautiously. Whichever is used, the clinician must be prudent as the models present a high risk of bias, according to PROBAST. The remaining models^{45-53,57,58} were not, as far as we are aware, validated externally, though they all obtained a discrimination (area under the ROC curve) above 0.70 (maximum 0.955, with values generally around 0.75). Accordingly, if these models were validated externally with good results, their use could also be considered, again with caution.

CONCLUSIONS

The mortality prediction models analyzed in this systematic review showed great variability in their designs and methodologies. After applying PROBAST as the gold standard to verify whether they were correctly developed, we found that only one study obtained a positive score in applicability together with an unclear risk of bias. The rest of the studies were found to have a high risk. Moreover, only three of the studies examined were validated externally, two just once but the other on four occasions.





REFERENCES

1. Galeski DF, Edwards JM, Kallan MJ, et al. Benchmarking the incidence and mortality of severe sepsis in the United States. *Crit Care Med* 2013; **41**: 1167-1174.
2. Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Med* 2017;**43**:304-377.
3. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016;**315**:801-810.
4. Hemingway H, Croft P, Perel P, et al; PROGRESS Group. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;**346**:e5595.
5. Steyerberg EW, Moons KG, van der Windt DA, et al; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;**10**:e1001381.
6. Riley RD, Hayden JA, Steyerberg EW, et al; PROGRESS Group. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 2013;**10**:e1001380.
7. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 2011;**9**:103.
8. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 2015;**162**:55-63.
9. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014;**11**:e1001744.
10. Wolff RF, Moons KGM, Riley RD, et al; PROBAST Group†. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;**170**:51-58.

11. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019;**170**:W1-W33.
12. Higgins JP, Altman DG, Gøtzsche PC, et al; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928.
13. Palazón-Bru A, Mares-García E, López-Bru D, et al. A critical appraisal of the clinical applicability and risk of bias of the predictive models for mortality and recurrence in patients with oropharyngeal cancer: Systematic review. *Head Neck* 2019. doi: 10.1002/hed.26025.
14. Palazón-Bru A, Mares-García E, López-Bru D, et al. A systematic review of predictive models for recurrence and mortality in patients with tongue cancer. *Eur J Cancer Care (Engl)* 2019;**28**:e13157.
15. Ueda T, Aoyama-Ishikawa M, Nakao A, Yamada T, Usami M, Kotani J. A simple scoring system based on neutrophil count in sepsis patients. *Med Hypotheses* 2014;**82**:382-386.
16. Mikacenic C, Price BL, Harju-Baker S, et al. A Two-Biomarker Model Predicts Mortality in the Critically Ill with Sepsis. *Am J Respir Crit Care Med* 2017;**196**:1004-1011.
17. Moreno RP, Metnitz B, Adler L, Hoechstl A, Bauer P, Metnitz PG; SAPS 3 Investigators. Sepsis mortality prediction based on predisposition, infection and response. *Intensive Care Med* 2008;**34**:496-504.
18. Howell MD, Talmor D, Schuetz P, Hunziker S, Jones AE, Shapiro NI. Proof of principle: the predisposition, infection, response, organ failure sepsis staging system. *Crit Care Med* 2011;**39**:322-327.
19. Shapiro NI, Wolfe RE, Moore RB, Smith E, Burdick E, Bates DW. Mortality in Emergency Department Sepsis (MEDS) score: a prospectively derived and validated clinical prediction rule. *Crit Care Med* 2003;**31**:670-675.

20. Chen KF, Liu SH, Li CH, et al. Development and validation of a parsimonious and pragmatic CHARM score to predict mortality in patients with suspected sepsis. *Am J Emerg Med* 2017;**35**:640-646.
21. Fortis S, O'Shea AMJ, Beck BF, et al. An automated computerized critical illness severity scoring system derived from APACHE III: modified APACHE. *J Crit Care* 2018;**48**:237-242.
22. Jaimes F, Farbiarz J, Alvarez D, Martínez C. Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. *Crit Care* 2005;**9**:R150-6.
23. Shapiro NI, Howell MD, Talmor D, Donnino M, Ngo L, Bates DW. Mortality in Emergency Department Sepsis (MEDS) score predicts 1-year mortality. *Crit Care Med* 2007;**35**:192-198.
24. Sivayoham N, Blake LA, Tharimooopantavida SE, et al. The REDS score: a new scoring system to risk-stratify emergency department suspected sepsis: a derivation and validation study. *BMJ Open* 2019 **26**;9:e030922.
25. Ward L, Paul M, Andreassen S. Automatic learning of mortality in a CPN model of the systemic inflammatory response syndrome. *Math Biosci* 2017;**284**:12-20.
26. Zhang Z, Hong Y. Development of a novel score for the prediction of hospital mortality in patients with severe sepsis: the use of electronic healthcare records with LASSO regression. *Oncotarget* 2017;**8**:49637-49645.
27. Garcia-Simon M, Morales JM, Modesto-Alapont V, et al. Prognosis Biomarkers of Severe Sepsis and Septic Shock by 1H NMR Urine Metabolomics in the Intensive Care Unit. *PLoS One* 2015 **13**;10:e0140993.
28. Rubulotta F, Marshall JC, Ramsay G, Nelson D, Levy M, Williams M. Predisposition, insult/infection, response, and organ dysfunction: A new model for staging severe sepsis. *Crit Care Med* 2009;**37**:1329-1335.

29. Santos TM, Franci D, Coutinho CM, et al. A simplified ultrasound-based edema score to assess lung injury and clinical severity in septic patients. *Am J Emerg Med* 2013;**31**:1656-1660.
30. Saka G, Kreke JE, Schaefer AJ, Chang CC, Roberts MS, Angus DC; GenIMS Investigators. Use of dynamic microsimulation to predict disease progression in patients with pneumonia-related sepsis. *Crit Care* 2007;**11**:R65.
31. Dumas G, Lavillegrand JR, Joffre J, et al. Mottling score is a strong predictor of 14-day mortality in septic patients whatever vasopressor doses and other tissue perfusion parameters. *Crit Care* 2019;**23**:211.
32. Asiiimwe SB, Abdallah A, Ssekitoleko R. A simple prognostic index based on admission vital signs data among patients with sepsis in a resource-limited setting. *Crit Care* 2015;**19**:86.
33. Baumgartner JD, Büla C, Vaney C, Wu MM, Eggimann P, Perret C. A novel score for predicting the mortality of septic shock patients. *Crit Care Med* 1992;**20**:953-960.
34. Guillén J, Jiankun L, Furr M, et al. Predictive Models for Severe Sepsis in Adult ICU Patients. 2015 Systems and Information Engineering Design Symposium 2015: 182-187doi: 10.1109/SIEDS.2015.7116970.
35. Slotman GJ. Prospectively validated prediction of physiologic variables and organ failure in septic patients: The Systemic Mediator Associated Response Test (SMART). *Crit Care Med* 2002;**30**:1035-1045.
36. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;**22**:707-710.
37. Hebert PC, Drummond AJ, Singer J, Bernard GR, Russell JA. A simple multiple system organ failure scoring system predicts mortality of patients who have sepsis syndrome. *Chest* 1993;**104**:230-235.

38. Lagu T, Rothberg MB, Nathanson BH, Steingrub JS, Lindenauer PK. Incorporating initial treatments improves performance of a mortality prediction model for patients with sepsis. *Pharmacoepidemiol Drug Saf* 2012;**21 (Suppl 2)**:44-52.
39. Sivayoham N, Holmes P, Cecconi M, Rhodes A. Internal emergency department validation of the simplified MISSED score. *Eur J Emerg Med* 2015;**22**:321-326.
40. Lissalde-Lavigne G, Combescure C, Muller L, et al. Simple coagulation tests improve survival prediction in patients with septic shock. *J Thromb Haemost* 2008;**6**:645-653.
41. Kim YC, Song JE, Kim EJ, et al. A Simple Scoring System Using the Red Blood Cell Distribution Width, Delta Neutrophil Index, and Platelet Count to Predict Mortality in Patients With Severe Sepsis and Septic Shock. *J Intensive Care Med* 2019;**34**:133-139.
42. Perl TM, Dvorak L, Hwang T, Wenzel RP. Long-term survival and function after suspected gram-negative sepsis. *JAMA* 1995 26;**274**:338-345.
43. West TE, Wikraiphat C, Tandhavanant S, et al. Patient Characteristics, Management, and Predictors of Outcome from Severe Community-Onset Staphylococcal Sepsis in Northeast Thailand: A Prospective Multicenter Study. *Am J Trop Med Hyg* 2017;**96**:1042-1049.
44. Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of In-hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data-Driven, Machine Learning Approach. *Acad Emerg Med* 2016;**23**:269-278.
45. García-Lamberechts EJ, Martín-Sánchez FJ, Julián-Jiménez A, et al. Infection and systemic inflammatory response syndrome in older patients in the emergency department: a 30-day risk model. *Emergencias* 2018;**30**:241-246.
46. Jiang W, Zhong W, Deng Y, et al. Evaluation of a combination "lymphocyte apoptosis model" to predict survival of sepsis patients in an intensive care unit. *BMC Anesthesiol* 2018 18;**18**:89.

47. Samsudin MI, Liu N, Prabhakar SM, et al. A novel heart rate variability based risk prediction model for septic patients presenting to the emergency department. *Medicine (Baltimore)* 2018;**97**:e10866.
48. Schwarzkopf D, Fleischmann-Struzek C, Rüdell H, Reinhart K, Thomas-Rüdell DO. A risk-model for hospital mortality among patients with severe sepsis or septic shock based on German national administrative claims data. *PLoS One* 2018 20;**13**:e0194371.
49. Bewersdorf JP, Hautmann O, Kofink D, Abdul Khalil A, Zainal Abidin I, Loch A. The SPEED (sepsis patient evaluation in the emergency department) score: a risk stratification and outcome prediction tool. *Eur J Emerg Med* 2017;**24**:170-175.
50. Fang WF, Douglas IS, Chen YM, et al. Development and validation of immune dysfunction score to predict 28-day mortality of sepsis patients. *PLoS One* 2017 26;**12**:e0187088.
51. Ford DW, Goodwin AJ, Simpson AN, Johnson E, Nadig N, Simpson KN. A Severe Sepsis Mortality Prediction Model and Score for Use With Administrative Data. *Crit Care Med* 2016;**44**:319-327.
52. Seo MH, Choa M, You JS, et al. Hypoalbuminemia, Low Base Excess Values, and Tachypnea Predict 28-Day Mortality in Severe Sepsis and Septic Shock Patients in the Emergency Department. *Yonsei Med J* 2016;**57**:1361-1369.
53. Mohan A, Shrestha P, Guleria R, Pandey RM, Wig N. Development of a mortality prediction formula due to sepsis/severe sepsis in a medical intensive care unit. *Lung India* 2015;**32**:313-319.
54. Osborn TM, Phillips G, Lemeshow S, et al. Sepsis severity score: an internationally derived scoring system from the surviving sepsis campaign database*. *Crit Care Med* 2014;**42**:1969-1976.
55. Sivayoham N, Rhodes A, Cecconi M. The MISSED score, a new scoring system to predict Mortality In Severe Sepsis in the Emergency Department: a derivation and validation study. *Eur J Emerg Med* 2014;**21**:30-36.

56. Granja C, Póvoa P, Lobo C, Teixeira-Pinto A, Carneiro A, Costa-Pereira A. The predisposition, infection, response and organ failure (Piro) sepsis classification system: results of hospital mortality using a novel concept and methodological approach. *PLoS One* 2013;**8**:e53885.
57. Lagu T, Lindenauer PK, Rothberg MB, et al. Development and validation of a model that uses enhanced administrative data to predict mortality in patients with sepsis. *Crit Care Med* 2011;**39**:2425-2430.
58. Tabak YP, Johannes RS, Silber JH. Using automated clinical data for risk adjustment: development and validation of six disease-specific mortality predictive models for pay-for-performance. *Med Care* 2007;**45**:789-805.
59. Du M, Haag D, Song Y, Lynch J, Mittinty M. Examining Bias and Reporting in Oral Health Prediction Modeling Studies. *J Dent Res* 2020. doi: 10.1177/0022034520903725.
60. Di Tanna GL, Wirtz H, Burrows KL, Globe G. Evaluating risk prediction models for adults with heart failure: A systematic literature review. *PLoS One* 2020;**15**:e0224135.
61. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ* 2019;**367**:l5358.
62. Carrillo-Larco RM, Aparcana-Granda DJ, Mejia JR, Barengo NC, Bernabe-Ortiz A. Risk scores for type 2 diabetes mellitus in Latin America: a systematic review of population-based studies. *Diabet Med* 2019;**36**:1573-1584.
63. Carrillo-Larco RM, Altez-Fernandez C, Pacheco-Barrios N, et al. Cardiovascular Disease Prognostic Models in Latin America and the Caribbean: A Systematic Review. *Glob Heart* 2019;**14**:81-93.
64. Weřna M, Adamik B, Goździk W, Kübler A. External validation of the sepsis severity score. *Int J Immunopathol Pharmacol* 2020;**34**:2058738420936386.

65. Sathaporn N, Khwannimit B. Validation the performance of New York Sepsis Severity Score compared with Sepsis Severity Score in predicting hospital mortality among sepsis patients. *J Crit Care* 2019;**53**:155-161.
66. Khwannimit B, Bhurayanontachai R, Vattanavanit V. Validation of the Sepsis Severity Score Compared with Updated Severity Scores in Predicting Hospital Mortality in Sepsis Patients. *Shock* 2017;**47**:720-725.
67. Williams JM, Greenslade JH, Chu K, Brown AF, Lipman J. Severity Scores in Emergency Department Patients With Presumed Infection: A Prospective Validation Study. *Crit Care Med* 2016;**44**:539-547.
68. Caramello V, Macciotta A, Beux V, De Salve AV, Ricceri F, Boccuzzi A. Validation of the Predisposition Infection Response Organ (PIRO) dysfunction score for the prognostic stratification of patients with sepsis in the Emergency Department. *Med Intensiva* 2020. doi:10.1016/j.medin.2020.04.022.
69. Ryoo SM, Ahn S, Kim WY, Lim KS. External validation of the MISSED score to predict mortality in patients with severe sepsis and septic shock in the emergency department. *Eur J Emerg Med* 2015;**22**:327-330.

FIGURE LEGENDS

Figure 1: PRISMA flow chart for the systematic review of prognostic models in sepsis.

PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses.





Table 1: Discrimination and calibration of the included predictive models in our systematic review.

References	Discrimination (AUC with 95% CI)	Calibration
García-Lamberechts et al., 2018⁴⁵	0.79 (0.72-0.85)	$p=0.496$ (Hosmer-Lemeshow test)
Jiang et al., 2018⁴⁶	0.955 (0.901-1)	None
Samsudin et al., 2018⁴⁷	0.79 (0.72-0.86)	None
Schwarzkopf et al., 2018⁴⁸	0.737 (0.734-0.74) (derivation) 0.736 (0.733-0.739) (validation) 0.739 (0.737-0.743) (validation)	Mortality rate in lowest risk-decile and highest decile (derivation): 0.11-0.78 Mortality rate in lowest risk-decile and highest decile (validation): 0.1-0.76 Mortality rate in lowest risk-decile and highest decile (validation): 0.1-0.77 Calibration plots (expected and observed risk were quite similar)
Bewersdorf et al., 2017⁴⁹	0.83 (0.79-0.88) (derivation) 0.80 (0.73-0.90) (validation)	$p=0.42$ (Hosmer-Lemeshow test for the development data set)
Fang et al., 2017⁵⁰	0.853 (0.769-0.915) (derivation) 0.789 (0.641-0.896) (validation)	None
Ford et al., 2016⁵¹	0.801 (derivation) 0.765 (validation)	$p=0.37$ (Hosmer-Lemeshow test for the development data set) $p=0.56$ (Hosmer-Lemeshow test for the validation data set) Calibration plots (expected and observed risk were quite similar)
Seo et al., 2016⁵²	0.8173 (0.7605–0.8741) (derivation) 0.7537 (0.6563–0.8512) (validation)	$p=0.61$ (Hosmer-Lemeshow test for the development data set) Calibration plots (expected and observed risk were quite similar)
Mohan et al., 2015⁵³	0.8422	Not given
Osborn et al., 2014⁵⁴	0.736 (derivation) 0.748 (validation)	$p=0.373$ (Hosmer-Lemeshow test for the development data set) $p=0.937$ (Hosmer-Lemeshow test for the validation data set) Calibration plots (expected and observed risk were quite similar)
Sivayoham et al., 2013⁵⁵	0.712 (0.64-0.78) (derivation) 0.68 (0.6-0.77) (validation)	$p=0.6$ (Hosmer-Lemeshow test for the development data set) $p=0.3$ (Hosmer-Lemeshow test for the validation data set)
Granja et al., 2013⁵⁶	0.66 (0.62-0.69) (Model 1) 0.59 (0.55-0.63) (Model 2) 0.72 (0.68-0.75) (Model 3) 0.81 (0.78-0.84) (Model 4) 0.84 (0.81-0.87) (Model 5)	$p=0.297$ (Hosmer-Lemeshow test for the Model 1) $p=0.630$ (Hosmer-Lemeshow test for the Model 2) $p=0.678$ (Hosmer-Lemeshow test for the Model 3) $p=0.588$ (Hosmer-Lemeshow test for the Model 4) $p=0.368$ (Hosmer-Lemeshow test for the Model 5)
Lagu et al., 2011⁵⁷	0.78 (derivation) 0.693 (0.649-0.775) (validation)	Calibration plot for derivation (expected and observed risk were quite similar) $p=0.11$ (Hosmer-Lemeshow test for the validation data set)
Tabak et al., 2007⁵⁸	Not given	Calibration plots (expected and observed risk were quite similar)

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval. When the CI is not given in the original paper, we have not been able to include it in the table.

Table 2: Risk of bias and concern regarding applicability of the included studies that evaluated multivariable prognostic prediction models for mortality in patients with sepsis (PROBAST).

Study	ROB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
García-Lamberechts et al., 2018 ⁴⁵	+	+	+	-	+	+	+	-	+
Jiang et al., 2018 ⁴⁶	-	?	?	-	-	?	?	-	-
Samsudin et al., 2018 ⁴⁷	-	+	+	-	-	+	+	-	-
Schwarzkopf et al., 2018 ⁴⁸	+	+	+	?	+	+	+	?	+
Bewersdorff et al., 2017 ⁴⁹	+	+	+	-	+	+	+	-	+
Fang et al., 2017 ⁵⁰	+	?	?	-	+	?	?	-	?
Ford et al., 2016 ⁵¹	+	-	+	-	+	-	+	-	-
Seo et al., 2016 ⁵²	+	+	+	-	+	+	+	-	+
Mohan et al., 2015 ⁵³	-	-	+	-	-	-	+	-	-
Osborn et al., 2014 ⁵⁴	+	-	+	-	+	-	+	-	-
Sivayoham et al., 2014 ⁵⁵	+	+	+	-	+	+	+	-	+
Granja et al., 2013 ⁵⁶	+	-	+	-	+	-	+	-	-
Lagu et al., 2011 ⁵⁷	?	?	+	-	?	?	+	-	?
Tabak et al., 2007 ⁵⁸	+	+	+	-	+	+	+	-	+

Abbreviations: PROBAST, Prediction model Risk Of Bias Assessment Tool; ROB, risk of bias.

+: low ROB/low concern regarding applicability; -, high ROB/high concern regarding applicability; ?, unclear ROB/unclear concern regarding applicability.

Table 3: Low risk of bias and low concern regarding applicability (%) of other systematic reviews that evaluated multivariable prediction models for other diseases (PROBAST).

Study	Disease	ROB				Applicability			Overall	
		Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Our review	Sepsis	71.4	50	85.7	0	71.4	50	85.7	0	42.9
Palazón-Bru et al., 2020 ⁵³	Oropharyngeal cancer	83.3	100	100	0	83.3	100	100	0	83.3
Du et al., 2020 ⁵⁹	Oral health	33.3	79.2	70.8	8.3	N/A	N/A	N/A	4.2	N/A
Di Tanna et al., 2020 ⁶⁰	Heart failure	75.0	62.5	65.0	20.0	80.0	97.5	82.5	17.5	87.5
Bellou et al., 2019 ⁶¹	COPD	82.8	96.6	98.5	2.5	N/A	N/A	N/A	1.7	N/A
Carrillo-Larco et al., 2019 ⁶²	Diabetes mellitus	83.3	50	50	0	83.3	83.3	83.3	0	83.3
Carrillo-Larco et al., 2019 ⁶³	Cardiovascular diseases	100	100	87.5	0	50	100	100	0	50

Abbreviations: COPD, Chronic Obstructive Pulmonary Disease; N/A, not assessed; PROBAST, Prediction model Risk Of Bias Assessment Tool; ROB, risk of bias.

APPENDICES

Tables E1-5: Relevant items extracted from the included studies based on the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies checklist.

Appendix E1: Search equations of the systematic review.

Appendix E2: Assessment of prognostic model development and validation studies for mortality in patients with sepsis.





14. AGRADECIMIENTOS





Especialmente a Jose, África y Lucas. Jose, has sido siempre mi gran apoyo. A África y Lucas porque os habéis convertido en el centro y motor de mi vida.

A mis hermanos, por haber sido mi referencia y mi guía.

A todos los miembros de mi familia que han confiado en mí y que me han apoyado incondicionalmente siempre.

A todos mis amigos y amigas que son un pilar importante en mi vida.





Mis agradecimientos a:

Antonio Palazón porque sin tu ayuda este proyecto no hubiera sido posible, metódico y trabajador, haces fácil cualquier tarea por ardua que sea. Ha sido un placer conocerte y trabajar a tu lado.

M. Ángeles Carbonell por tu motivación y ejemplo de perseverancia, responsabilidad y compromiso. Un trabajo a tu lado es sinónimo de superación y éxito.

Vicente Gil por confiar en mí para este proyecto, figura imprescindible para la consecución del mismo. Gracias por tu presencia en los momentos fundamentales.

Patricia Llópez, amiga y compañera, luchadora incansable, gracias por compartir y contagiar el buen ánimo que te caracteriza, trabajar cada día a tu lado es un lujo. Juntas hemos hecho realidad este proyecto.

A mis compañeros del Hospital de Elda, gracias por el equipo humano que sois. A mis compañeros del Hospital de Almansa, os recordaré siempre.



“Desde que tengo memoria, siempre lo he querido todo de la vida, todo lo que pudiera concederme, un deseo que me aparta de quienes están dispuestos a conformarse con menos. No entiendo cómo se puede desear menos, cómo se puede albergar ambiciones limitadas y ridículas cuando las posibilidades son infinitas”

“Me he liberado de mi pasado, pero no lo he anulado. Valoro los momentos y las experiencias que me dieron forma. He vivido la historia. La gente quiere saber si he encontrado la felicidad, pero lo que he encontrado es mejor aún: la autenticidad. Por fin soy libre para ser yo misma, y eso siento muy bien. Si alguna vez alguien intenta decirte que seas algo que no eres, espero que también tú encuentres el valor para levantar la voz y protestar”.

Deborah Feldman. Unorthodox.



